Johann Ari Larusson · Brandon White
Editors

Learning Analytics

From Research to Practice

Springer

Learning Analytics

Johann Ari Larusson • Brandon White
Editors

Learning Analytics

From Research to Practice

Springer

Editors
Johann Ari Larusson
Center for Digital Data, Analytics
and Adaptive Learning at Pearson
Boston, Massachusetts, USA

Brandon White
Department of English
University of California, Berkeley
Berkeley, California, USA

ISBN 978-1-4614-3304-0 ISBN 978-1-4614-3305-7 (eBook)
DOI 10.1007/978-1-4614-3305-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014940734

© Springer Science+Business Media New York 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To our families

Acknowledgements

As editors with interests in pedagogy, community, and collaboration, it is only fitting that we begin by expressing our gratitude to our authors, each of whom has made the production of this volume such a valuable learning experience in its own right. Thanks to their efforts, we are confident that this book will make a forceful first contribution to the field of learning analytics. In particular, we would also like to thank Ryan Baker and Stephanie Teasley for their separate early insights into the project's potential merits, scope, and organization. We have kept these suggestions in mind since the first, and they have been instrumental in shaping the volume that we hold today.

We would moreover like to thank the organizers and members of the Society for Learning Analytics Research (SOLAR) for their foundational efforts in providing a formal and quickly evolving home for the study of learning analytics. It was through the collective achievements and enthusiasms of the participants at the 2011 conference on Learning Analytics and Knowledge (before SOLAR even had that name) that the prospect of such a collection first came together.

This book could not have been assembled without the efforts of the many people on Springer's editorial and production staff, specifically Melissa James and Miriam Kamil. We would moreover like to thank those who helped us with our own side of the editorial and production process, Mike Vetter and Diego Villamil.

Finally, we want to extend our utmost gratitude to our colleagues at Brandeis University, Pearson, and the University of California, Berkeley for their support and guidance throughout this process. Your every effort has helped to reinforce what a learning community, at its best, might be.

Johann Ari Larusson
Brandon White

Contents

Contributors

Ryan Shaun Baker is Associate Professor of Cognitive Studies in Education in the Department of Human Development at Teachers College Columbia University and Assistant Professor of Social Science and Policy Studies at Worcester Polytechnic Institute.

John T. Behrens is Vice President of Learning Analytics for the Pearson School Group and Vice President of the Center for Digital Data, Analytics and Adaptive Learning at Pearson. He is also an Adjunct Assistant Research Professor in the Department of Psychology at the University of Notre Dame.

Christopher Brooks is a Research Fellow at the School of Information, University of Michigan.

John P. Campbell is Associate Provost and Chief Information Officer at West Virginia University, Morgantown, West Virginia.

Kristen E. DiCerbo is a Principal Research Scientist in the Center for Digital Data, Analytics and Adaptive Learning at Pearson and a Distinguished Learning Game Researcher at GLASSLAB.

Jim Greer is a Professor of Computer Science and Director of the University Learning Centre at the University of Saskatchewan.

Carl Gutwin is a Professor of Computer Science at the University of Saskatchewan, with a research focus on Human Computer Interaction (HCI) and Computer Supported Collaborative Work (CSCW).

Paul Salvador Inventado is Assistant Professor in the Center for Empathic Human-Computer Interactions, College of Computer Studies, De La Salle University, Philippines and a Ph.D. student in Architecture for Intelligence in The Institute of Scientific and Industrial Research, Osaka University, Japan.

Andrew E. Krumm is an Education Researcher at SRI International, Center for Technology in Learning.

Johann Ari Larusson is a Senior Research Scientist for the Center for Digital Data, Analytics and Adaptive Learning at Pearson, Boston, Massachusetts, USA.

Steven Lonn is Assistant Director of USE Lab and Library Learning Analytics Specialist at the University of Michigan.

Abelardo Pardo is Lecturer in the School of Electrical and Information Engineering, University of Sydney, Australia.

Matthew D. Pistilli is a Research Scientist for Teaching and Learning Technologies for the Instruction Design Center within Information Technology at Purdue University, West Lafayette, Indiana.

Stephanie D. Teasley is Research Professor in the School of Information and USE Lab Director at the University of Michigan.

R. Joseph Waddington is a Postdoctoral Research Associate at the University of Notre Dame, Institute for Educational Initiatives.

Brandon White is a doctoral candidate in English at the University of California, Berkeley.

James E. Willis III is an Educational Assessment Specialist for the Instructional Design Center within Information Technology at Purdue University, West Lafayette, Indiana.

Chapter 1
Introduction

Johann Ari Larusson and Brandon White

Supporting teaching and learning with technology is becoming as commonplace as chalk in today's educational institutions. However, simply making technology available or requiring students to use it does not necessarily guarantee success. How does one effectively explore online learning communities so as to get an accurate description of the complex interactions taking place? What methods for analysis are available? What does a method of analysis even look like? What is the unit of analysis? How can an institution effectively organize its data? How does the information collected enrich students' learning experiences? How can we positively impact the teachers' pedagogical practices? How does one even design for successful implementations of educational technology that report back data rich enough to affect subsequent implementations? These questions are the ones required to better inform an educational agenda not only for "teaching with technology," but simply for teaching in the first place.

These processes, whatever their form, are inherently complex. The teaching and learning themselves might be taking place in a classroom, but are all nevertheless unfolding in an intangible time and space—inside a "black box," so to speak—producing enormous volumes of data where the vision of what data to collect, how to collect it, and how to explore it is not necessarily clear. In recent years, learning analytics (LA) has emerged as a field that seeks to provide answers to questions such as the ones highlighted above. Learning analytics can be summarized as the collection, analysis, and application of data accumulated to assess the behavior of educational communities. Whether it be through the use of statistical techniques and predictive

J.A. Larusson (✉)
Center for Digital Data, Analytics and Adaptive Learning at Pearson,
Boston, Massachusetts, USA
e-mail: johann.larusson@pearson.com

B. White
Department of English, University of California, Berkeley, Berkeley, California, USA
e-mail: brandonw@berkeley.edu

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_1, © Springer Science+Business Media New York 2014

modeling, interactive visualizations, or taxonomies and frameworks, the ultimate goal is to optimize both student and faculty performance, to refine pedagogical strategies, to streamline institutional costs, to determine students' engagement with the course material, to highlight potentially struggling students (and to alter pedagogy accordingly) to fine-tune grading systems using real-time analysis, and to allow instructors to judge their own educational efficacy. In every case, learning analytics gives all stakeholders insight into what is taking place from Day 1 to Day X of a given class irrespective of the type of activity taking place. In short, learning analytics is broadly defined as the effort to improve teaching and learning through the targeted analysis of student demographic and performance data (Elias 2011; Fritz 2010). The contents of the "black box," in other words, become that much more visible, with their various markers sampled, collected, evaluated, and replayed in a legible form.

Learning analytics encompasses a range of cutting-edge educational technologies, methods, models, techniques, algorithms, and best practices that provide all members of an institution's community with a window into what actually takes place over the trajectory of a student's learning. Involvement in LA technologies and pedagogies allows educators and scholars to engage in a contemporary and innovative approach to an educational issue that is already an integral part of higher education.

In many ways, the field of learning analytics should be considered new. The field itself has come into being largely thanks to the proliferation of digital data produced by educational institutions' increasing tendency to produce, submit, and assess academic work in electronic form (Greer and Heaney 2004; Hirst 2011). While the first formal conference on LA, held in 2011, is evidence of its growing relevance in educational circles on an international scale, the fact that such a conference had not existed previously is sign enough of LA's relative infancy.

Learning analytics ideally attempts to leverage data to provide insight into the activities taking place within the classroom. What metrics are derived can then be fed back into pedagogy or applied with consequences even well outside the classroom itself. Several higher education institutions in particular have begun applying learning analytics to evaluate crucial aspects of the learning process and pedagogical practice, alongside institutional aims like student retention and cost reduction (Siemens and Long 2011). Holistic descriptions of several of these practices can be found in (Siemens and Long 2011) and (Ferguson 2012). A recent U.S. Department of Education brief held that learning analytics prioritizes the "*human tailoring* of responses, such as through adapting instruction content, intervening with at-risk students, and providing feedback" (Bienkowski et al. 2012, p. 13). This approach "does not emphasize reducing learning into components but instead seeks to understand entire systems and to support human decision making" (ibid). Yet for all the budding interest in LA, its earliest implementations have evolved from older models and methods, from raw data mining (cf. Baker and Yacef 2009) and learning community studies (cf. Dawson 2010) to the broader field of academic analytics (Goldstein and Katz 2005; Campbell et al. 2006). As institutions and educators increasingly begin to install learning analytics systems, or learning analytics enabled systems, they often tend to employ frameworks inherited from several of these other fields. Even nominal attempts to directly improve learning and teaching practice tend to digest institutional systems data with limited understanding of how that data

could or should inform pedagogy. Although these other inquiries remain vital and valuable fields, the purpose of this volume is to help situate LA's unique priorities, unique intended benefits, and unique ranges of personnel capable of putting that technology into practice. Growing the field of learning analytics requires making sure that it remains distinct from what came before and that its purpose remains rigorously clear.

Up until recent years, research and practice in this area has been hampered by a lack of definition, with work in the field dispersed throughout a number of journals and conferences, making it more difficult for experts to share results, get a real sense of what is new and innovative, or to identify the best practices, strategies, or tools to use. The birth of learning analytics as a field of study in its own right, through a now annual conference, the recently established Society for Learning Analytics Research, and with workshops and symposiums being organized around the world, has now made it possible to consolidate research once taking place along its periphery under one umbrella.

Learning analytics is uniquely positioned as a field with the potential to guide the efforts of any of a number of institutional actors or stakeholders, from students to instructors, IT professionals to educational administrators. While the inputs of learning analytics derive primarily from the classroom, any one of these stakeholders may well be charged with evaluating the results, putting changes into action, and weighing the impact that results. This book attempts to provide the first comprehensive reference book for LA, with the aim of helping scholars, researchers, developers, IT professionals, chief technology and information officers, university administrators, or anyone and everyone interested in advancing the field of learning analytics by showcasing the latest results, strategies, guidelines, methods, models, and tools. Collecting all of this information in one volume will allow scholars and researchers to take stock of ongoing efforts in the field, helping to illuminate what areas remain to be explored, and thus pushing the field yet further forward.

The purpose of this volume is, simply put, to provide an entry point into the field for any one of these actors depending upon their unique institutional interests. As a field with a broad appeal, simply navigating the extant literature of learning analytics, let alone attempting to put any of those principles into practice, can prove daunting. The chapters that follow each attempt to consolidate much of the available literature while putting forth best practice guidelines or model case studies that might prove of interest to particular types of readers. As such, this book is organized not around common problems or the mounting complexity of its efforts, but rather around the kinds of communities that each chapter attempts to address. It is the hope of the editors that this approach will allow different kinds of readers an opportunity to easily identify those chapters most likely to offer immediate insights. In the remainder of this introduction, each chapter's possible contributions to the field are thus suggested alongside discussion of its possible appeal to different classes of readers. Rather than simply summarizing what follows, the reader can think of this chapter as a map to the different ways in which the book itself might be read.

These myriad types of engagement are what the complexity of the field of learning analytics requires. What is learning analytics? The answer is not simply our own, as editors, but the one that the book itself, through each of our various

contributors, comes to suggest. These inquiries *themselves are* learning analytics. By probing the field's theoretical investments and by classifying its possible components, by exploring its history relative to educational data mining (EDM) and by helping to place it within a broader institutional context, by applying case studies to educators and students and academic advisors alike, these chapters take stock of all the many stakeholders that learning analytics might attempt to benefit, and thus most comprehensively demonstrate its power and its promise.

1.1 Preparing for Learning Analytics

The first section of the book, "Preparing for Learning Analytics," looks to clarify the stakes of learning analytics by supplying suggestions for the field's domain, potential, and possible points of emphasis. In each of the three chapters, these suggestions take the form of guidelines for what the development of a learning analytics application might require. Whereas later chapters begin with established technologies or established pilot programs at host universities, these chapters take none of that for granted, investigating instead the very foundations of learning analytics practice.

An initial entry point for nearly any reader can be found in Chap. 2, Abelardo Pardo's "Designing Learning Analytics Experiences." Pardo synthesizes the results and proposals of dozens of research findings to suggest five phases of design and execution through which an LA intervention might pass. These phases form a flexible framework that might be applied to any LA endeavor, providing readers with a sense of the kinds of decisions, dependencies, and trade-offs that are encountered in taking an analysis tool from conceptualization to subsequent enhancement.

The first stage, "capture," corresponds to the earliest collection of student data. The second stage, "report," delivers that data to a specifically defined set of stakeholders. The third stage, "prediction," deploys any of a number of techniques to provide non-intuitive answers to frequently encountered educational questions, such as the likelihood of an individual student failing a course or failing to graduate altogether. The "act" stage that follows offers the possibility of issuing automated solutions or implementing manual ones that have the potential, ideally, to reverse the most dire consequences of the earlier prediction. In the final stage, "refinement," the efficacy of the resulting actions is assessed anew so that the long-term viability of the analysis can itself be modified as need be.

Each stage in this process is presented not just as a single phase, with a predefined beginning and end, but as intimately bound up with choices that might have been made in earlier stages. Taking for instance only the "report" phase, Pardo singles out LA systems aimed at different classes of stakeholders. A process designed to deliver data directly to instructors requires a different set of investments than one generating metrics for IT professionals. Rather, therefore, than prescribing specific guidelines for what every stage ought to entail, Pardo instead offers a series of

questions at each stage that might inform how an LA implementation might be successfully designed and executed.

In a way, Pardo's chapter can be considered as a heuristic for much of the work of this volume as a whole. All of the subsequent chapters offer some specific engagement with one or more of the steps that Pardo outlines, and each is inevitably a product of choices that must have been made in any particular stage. Certain chapters, like that of Ryan Baker and Paul Inventado (see Chap. 4), are concerned principally with a specific phase: in this case, the kinds of calculations that are typically embraced during the "prediction" stage by learning analytics and EDM communities, respectively. Two case study chapters, as offered by Andrew Krumm et al. (see Chap. 6) and Brandon White and Johann Ari Larusson (see Chap. 8) provide considerations of two separate phases of analysis, carried out over a number of years. It might be suggested that these chapters have specific implications for what Pardo calls "refinement." Every attempted "prediction," however, is inevitably dependent on the stages that must have come before, every "refinement" only as good as the various acts that have been executed along the way.

Readers of this volume might thus do well to begin with Chap. 2, and to think of its many insights when taking up any of the chapters that follow. Such schematization of the assumptions underlying any particular technology will only lead to better questions, more pointed questions, and thus more opportunities for further refinement. Getting more LA systems to this final stage can be considered one of this volume's explicit goals. As Pardo mentions, because of LA's only relatively recent adoption and expansion, few LA systems have graduated to the "refinement" stage of analysis. There simply is not much longitudinal data on how a system might go through several iterations. It is our hope as editors that the next generation of LA research will see yet further instances of Pardo's framework being brought full circle.

Chapter 3, "Harnessing the Currents of the Digital Ocean" by John T. Behrens and Kristen E. DiCerbo, extends the discussion of the previous chapter to address the abundance of electronic information that now characterizes many educational efforts. Behrens and DiCerbo contrast this "digital ocean" with what they call the "digital desert," the pre-digital environment of the late twentieth century where data was rare, expensive to obtain, and as such was only amenable to limited, if any, analytic applications. While the potential consequences of such a shift can be seen as an impetus for the instantiation of learning analytics as a field in the first place, Behrens and DiCerbo suggest that the prevalence of digital data requires perhaps a more fundamental reshaping that the field might yet need to undergo. In their account, the simple technological limitations of the "digital desert" were themselves responsible for the kinds of activities, like multiple choice quizzes, that were developed for analysis. These activities tended—and still often tend—to be presented in a fixed form, with static questions matched to fixed answers to measure the correct response. As we stand on the shore of the "digital ocean," however, these same standards for collection, evaluation, and dissemination needn't remain a constraint.

Behrens and DiCerbo argue that the potential of learning analytics lies in its ability to reconceptualize the educational space, allowing us to think of user activity as

an ever-modulating stream of inputs from which certain attributes can be observed over time rather than requiring any moment-by-moment measure of correctness. The consequences for this shift in worldview would thus alter not only the ways in which data is collected, but also the very way in which it is understood. If the current understanding of summative evaluations (like final exams) interspersed with formative exercises (like homework, quizzes, midterms, or papers) can be likened to an autopsy conducted following a series of routine checkups, a naturalistically embedded assessment by way of learning analytics might instead be compared to a heart monitor that regularly and automatically generates feedback on the conditions at hand. Rather than seeing the analytic interface as something that delivers content to students, this understanding would instead see learning analytics as something allowing students themselves to create, explore, and reinforce the conditions of their own learning. As in Chap. 2, this discussion would situate learning analytics as an embedded system of continued research and refinement.

This chapter takes as its explicit endeavor an attempt to help readers rethink some of the underlying assumptions regarding how data and data analysis might be structured in a computationally complex space. Readers looking to expand their sense of what learning analytics might attempt would benefit from using this chapter as a primer on the possibilities while also helping to root the field's ambit in a broader history of educational theory. The chapter itself concludes with a wealth of suggestions for future research, not strictly in the *applications* of learning analytics, but on the types of thinking and training that learning analytics might require of the researchers themselves. These suggestions are invaluable as a basis from which to evaluate the core assumptions of learning analytics, and in so doing further suggest the kinds of practices, principles, and applications that await on the horizon.

Despite the only recent consolidation of learning analytics as a specific field of inquiry, LA approaches and LA methods have not developed in a vacuum. Chapter 4, Ryan Baker and Paul Inventado's "Educational Data Mining and Learning Analytics," examines learning analytics relative to the abutting discipline of EDM. Baker and Inventado provide a historical contextualization for EDM's growth, both as a research community and as a specific field of scientific inquiry.

Their particular focus, however, lies in the identification of several key methods that EDM has traditionally deployed that are perhaps more foreign to LA research, although they needn't necessarily be. These methods fall into the broad categories of "prediction models," "structure discovery," "relationship mining," and "discovery with models." For each category, Baker and Inventado not only identify relevant applications of the method, but discuss how each method has historically been a part of EDM research, and to what extent it remains so to this day.

By exploring each category more closely, Baker and Inventado are able to provide broad contextualizations of what EDM-type analysis might attempt. Their chapter could, in this way, be considered as a complement to Abelardo Pardo's chapter on learning analytics design (see Chap. 2). Yet the two projects—and, consequently, two learning communities—needn't be entirely at cross-purposes. The diachronic exploration of EDM's evolution that Baker and Inventado present is useful on the one hand for illuminating the areas where learning analytics researchers

have pursued EDMs through different means. It is useful in turn, on the other hand, for suggesting areas of inquiry that LA has heretofore left mostly untapped.

Readers looking to answer a specific subset of research questions might do well to consult this chapter as a kind of guide to what other strategies might continue to augment LA research. Taking a cue from John T. Behrens, one of the other authors in this volume (see Chap. 3), Baker and Inventado note that learning analytics and EDM essentially have their names reversed: that while learning analytics tends to focus on educational outcomes, EDM is more often than not concerned with the immediate products of learning. What this chapter potentially suggests is that a long-term implementation of LA and EDM methods in concert would ultimately wind up informing one another intimately, with improved learning coming to ensure continually optimal educational outcomes. At the point in the near future when both fields have readily demonstrated enough success that they can be readily installed, run, and refined—at the point, in other words, where research becomes practice—the difference between the two becomes virtually indistinguishable.

1.2 Learning Analytics for Communities

The second section of this volume explores learning analytics that speak to the specific interests of learning communities beyond the immediate teacher–student relationship. These chapters ask what it means to conceive of learning analytics at a large scale, either by discussing the implications of learning analytics for institutions as a whole, or by empowering a different level of stakeholders to leverage analytic insights.

One of the common concerns of several of this volume's authors lies in the granularity—or specificity—of the reporting data that an LA system might produce (see, for instance, Chap. 2). Data must be specific enough that its insights are made intelligible, but general enough that the end user isn't overwhelmed by abundant detail. Many of the chapters in this volume, such as Chap. 8 or Chap. 6, describe technologies meant to be put in the hands of on-the-ground users, be they instructors or students advisors. In a more fully integrated LA landscape, however, one can easily imagine any number of classrooms interventions taking place side by side. As soon as decisions about LA use need to be made beyond the individual classroom, a different series of questions immediately need to be considered. Chapter 5, "Learning Analytics at an Institutional Level," by Matthew D. Pistilli, James E. Willis, III, and John P. Campbell, describes the way in which an institutional actor, such as an administrator, a technology officer, or a system administrator, might go about the process of implementing and overseeing an LA architecture.

Building off of Tinto's theory of student departure, Astin's theory of student involvement, and Chickering and Gamson's principles for good practice in undergraduate education, the authors suggest a framework for where an institutional attempt at LA might even be committed. The standard that the authors put forth is ultimately a measure of a student's place in his or her educational environment.

Learning communities inevitably extend well outside the classroom, and even factors as casual as a student's frequency of contact with an instructor or involvement in the extracurricular games taking place nearby can stimulate a student's investment in his or her educational institution, increasing the likelihood that he or she will remain enrolled, excel in classes, and work towards a degree. Institutions themselves are thus ideally positioned to leverage observations of these interrelated interactions through analytics. Such a model of analytical practice takes stock of a diverse array of factors, gathered from a variety of different interactions, and uses the data from these interactions to suggest altered approaches that might improve a student's comfort, confidence, and capability in his or her educational setting.

Several consequences emerge from this analysis, the first of which is suggested even by the use cases sketched above. Pistilli et al. foremost suggest a renovation in the ways in which institutions even come to think of analytics in the first place, urging institutions to take stock of ambient data based on existing interactions between students, faculty, and supporting staff rather than going out and creating data sets from the ground up. The second suggestion informs the way in which such data might ultimately be used. Policies governing the privacy of information collected and disseminated in such a context are alone an important consideration for any such implementation, especially considering the varying standards for how confidential information might be handled at different universities even within the same country, city, or state. But what the authors ultimately put forward is a means for an institution to consider the interests of every other stakeholder concerned. It is not only, for instance, that faculty need to be sensitive to how they deploy analytics in their interactions with students, but that they also need to remain cognizant of the ways in which students could actually be *discouraged* by the result, leading to a negative feedback loop which is far from any LA implementation's intended purpose.

Readers of this volume with a particular interest in institutional efficacy would do well to consult this chapter early as a baseline look at what the commitments of an institution in an LA context are or could be. What this chapter suggests is that the most stable sense of analytics' place in an institution's daily life can only be understood holistically, as an aggregate consideration of the benefits accrued to every one of its individual actors.

Chapter 6, Andrew E. Krumm, R. Joseph Waddington, Stephanie D. Teasley, and Steven Lonn's "A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering" reports on an ongoing case study working hand-in-hand with stakeholders to develop a system capable of informing academic advisors of students in need of additional support. It shouldn't escape notice that this study, while directly dependent on student data, is the only chapter of this volume that doesn't use the individual course instructor as the primary instigator of interventions. This configuration of stakeholders thus suggests one immediate application of the kind of discussion found in Chap. 5. The relevant stakeholders here, and the ones who the authors approached with considerations for the second phase of their study, are the academic advisors who more often than not function as gatekeepers between instructional and institutional requirements.

The authors thus come to extend the conception of what a pedagogical intervention, properly carried out, might be.

As an early warning system, the authors' project was designed to alert academic advisors to whether students were likely to require further encouragement based on a number of data markers culled from a learning management system: graded activities, frequency of log-ins, and relative contextualization of a student's performance based on that of his or her own peers. This method thus combined real-time data with longitudinal tracking, creating a kind of self-correcting system. Since the intended recipients of this information were not involved in the day-to-day work of instruction, what the study ultimately attempted to improve upon was its own necessary granularity as determined by the frequency of the reports that advisors would receive.

Two possible classes of readers might be most interested in this volume's case studies (Chaps. 6–8), but in this chapter in particular: those readers looking even for a specific sense of what an LA application might entail in the first place, and those readers who, having surveyed the other contents of the book, now want to test how other broader principles might be put into practice. Readers of Chap. 6 would especially benefit from considering the chapter alongside Chap. 2. The early warning system described and refined in Chap. 6 remains the only instance of what Pardo calls "refinement" carried to its utmost, with the results of live analysis coming to actively reinflect the production cycle of a subsequent iteration of the tool. As such, the study discussed here is of value not only for how it might model what a successful refinement entails, but for what it might suggest about the future of learning analytics, when certain systems have become an established enough part of educational practice that they can routinely produce results but also be routinely improved.

1.3 Learning Analytics for Teachers and Learners

The final section of this book details attempts to use analytics to explore the environment most familiar to academic practice: the classroom. The reliance of learning analytics on student data has been a consistent theme throughout this volume. These chapters describe ways in which that data might be deployed, allowing instructors to use analytics to reshape or refine pedagogy.

Chapter 7, by Christopher Brooks, Jim Greer, and Carl Gutwin, "The Data-Assisted Approach to Building Intelligent Technology Enhanced Learning Environments" serves as a possible bridge between more theoretically oriented material and the case studies that follow. As the authors describe, intelligent tutoring systems have often been deployed as a means of scaling educational materials to better suit student performance. Such systems, however, are possibly unwieldy, requiring not just one expert in a given discipline or domain—the individual that we would ordinarily think of as a course instructor—but a separate pedagogical expert to weigh the multiple possible responses to a learner's mistake, and yet another series of experts to build tiers of content suitable to any number of learners. Such a tutoring system, in other words, becomes magnitudes more labor intensive than the

simple intensive instruction that the intelligent tutoring system might have been designed to ease. The authors thus propose a "data-assisted approach" to intelligent tutoring systems that acknowledges the instructor's place in conducting classes and manually performing pedagogical interventions.

Although the authors roughly categorize classes of implementations that might be pursued using a data-assisted approach, the heart of the chapter lies in an enumeration of specific motivating scenarios that describe three different educational contexts, three different technologies, and three different types of data for which a data-assisted intelligent technology would be of immediate use. These scenarios can most productively be considered as case studies of three different applications in their own right. In each instance, the authors not only showcase the application's functionality, but provide a supplementary consideration for how each application can be deployed at a different level of granularity, with a different specific focus, or with a different degree of buy-in from users.

Readers of almost any specific interest would do well to consult this chapter as a kind of test case for the possible range of even a single type of LA engagement. Those readers new to learning analytics may want to consult this chapter in concert with Chap. 2 as a means of measuring LA's possible prerogatives against how those prerogatives are developed in practice. One of the chapter's many merits is the way in which it suggests that the many choices confronting a new LA implementation needn't be binary choices at all, but decisions that the *right* implementation can pursue in parallel.

Chapter 8, "Identifying Points for Pedagogical Intervention Based on Student Writing: Two Case Studies for the 'Point of Originality'," by Brandon White and Johann Ari Larusson, provides a second case study chapter. This chapter showcases a computational method and tool called the "Point of Originality," which measures a student's ability to put key course concepts into his or her own words as a course progresses. With the mounting trend in higher education towards larger and larger gateway courses, especially in the early phases of a student's academic career, the Point of Originality is proposed as a way to let instructors quickly diagnose which students are likely to be struggling, and to use that information to conduct specific pedagogical interventions. As in Chap. 6, the use case for the tool is not an uncommon one: the only required inputs are the kinds of regular, iterative writing activities (be it a blog, discussion board, or other written assignment) that many instructors already use. Once an instructor has input a series of course concepts—either a few terms that might be likely to come up on an approaching midterm, or else a string of every key term found on the course syllabus—a custom algorithm calculates how proximately related every word in every student's writing sample is to those terms.

The chapter follows two different proof of concept case studies, the second conducted as a larger-scale elaboration of the first. The earlier case study uses actual course data to imagine a not unconventional scenario: an instructor is preparing to distribute an assignment midway through a semester, and wants to know whether students will likely be equipped to answer the prompt. Using the terms of the prompt as the initial "query term" concepts, the Point of Originality tool weighs the degree to which students have been able to put those same concepts into their own words at

any point during the semester, and provides a graphical and numerical representation of which students are likely to succeed and which are likely to struggle. The second case study adapts these findings for an entirely different course environment, one with many more students and with a highly technical scientific subject matter. The movement between these two case studies suggests the application of a principle suggested, in different ways, by each of the chapters in this volume (and perhaps most pointedly so in Chap. 3)—that the integrity of analytics efforts come from their ability to be universalized, used as well in one course context as in another. In an effort to streamline the possible applicability of the Point of Originality yet further, this second case study even removes the requirement that an instructor limit his or her query terms to those related to a specific assignment: rather, this case study simply made use of more than a 100 terms relevant to the full course syllabus, offering instructors insight into how students were approaching the course material in the most general terms. Both case studies demonstrated a strong degree of correlation between the metrics produced by the tool and a student's eventual performance in the class. The suggestion is that use of the Point of Originality tool would have singled out the struggling students well in advance, and would have allowed an instructor to take action as need be, either by working with those students or even by looking at the results to determine *which* of the concepts in circulation met with the most difficulty.

As in Chap. 6, readers might well be interested in this chapter as a way of determining how the principles outlined elsewhere in this volume can be applied to the design of an actual learning analytics platform, or for simply determining what a learning analytics system looks like to begin with. The chapter itself provides a slightly different shift in inquiry from several of the other chapters by examining a different *type* of data set than used elsewhere. As such, the chapter might usefully be considered alongside like attempts at analysis, not as a competing method, but as a tool that might be used in concert with a handful of others.

It is the possible harmony between several of these methods that forms one of the final suggestions of this volume. As learning analytics matures as a field, there will come a point where diagnostic methods are liable to overlap, where one analytics tool might be used by instructors while another relays a different set of information to institutional actors. The several contributions of this volume are all ultimately cross-compatible. Learning analytics depends on particular methods, particular metrics, particular tools, but the only learning analytics *solution* might well be a holistic solution, one that speaks equally to the experiences of learners, educators, and administrators.

References

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: U.S. Department of Education Office of Educational Technology.

Campbell, J. P., Finnegan, C., & Collins, B. (2006). Academic analytics: Using the CMS as an early warning system. *WebCT impact conference*, Chicago, IL.
Dawson, S. (2010). Online forum discussion interactions as an indicator of student community. *Australasian Journal of Educational Technology, 22*(4), 495–510.
Elias, T. (2011). *Learning analytics: Definitions, processes and potential*. Unpublished Internal Whitepaper of Athabasca University, Canada. Retrieved from https://landing.athabascau.ca/mod/file/download.php?file_guid=43713
Ferguson, R. (2012). Learning analytics: Development, drivers, and challenges. *International Journal of Technology Enhanced Learning, 4*(5–6), 304–317.
Fritz, J. (2010). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *The Internet and Higher Education, 49*, 89–97.
Goldstein, P. J., & Katz, R. N. (2005). *Academic Analytics: The Uses of Management Information and Technology in Higher Education*. Boulder: EDUCAUSE.
Greer, L., & Heaney, P. J. (2004). Real-time analysis of student comprehension: An assessment of electronic student response technology in an introductory earth science course. *Journal of Geoscience Education, 52*, 345–351.
Hirst, T. (2011). Pragmatic analytics: Insight, representation, and structure. *Keynote presented at the First International Conference Learning Analytics and Knowledge*, Banff, AB.
Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 31–40.

Part I
Preparing for Learning Analytics

Chapter 2
Designing Learning Analytics Experiences

Abelardo Pardo

2.1 Introduction

The use of technology in learning experiences is an area that has experienced unprecedented growth over the years. The appearance of so-called Learning Management Systems (LMSs) allowed certain tasks related to learning to be technologically mediated. Other areas such as automatic assessment, discussion forums, or tutoring were also disrupted by emerging technology. The use of such technology in today's educational institutions varies significantly depending on the age of the students and the geographical location of the institution, and thus, its impact is equally varied. But one common trend that can be considered valid across all these contexts is that the barrier to access electronic information produced by and required for learning experiences has lowered if not disappeared. This fact has led to a significant change in the forces that typically shape how a learning experience is designed and deployed.

The current trend is characterized by the increase of technology mediation in aspects related to a learning experience. This mediation allows tools to collect a wide variety of observations without user intervention. Up until now, collecting evidence, events, opinions, or generic feedback explicitly from the students and using it as feedback to improve a learning experience had a fairly reduced impact. A detailed observation of how students learn had previously required invasive methods that would interfere with the actual teaching and learning process. Nowadays, with the ubiquitous presence of digital devices mediating interactions, technology offers the possibility of collecting a comprehensive set of observations of the interactions occurring in a learning environment with almost no user intervention. Analytic techniques already used in areas such as business intelligence are now

A. Pardo (✉)
School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia
e-mail: abelardo.pardo@sydney.edu.au

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_2, © Springer Science+Business Media New York 2014

applied in a learning experience in what is called *learning analytics* (Long and Siemens 2011). Learning Analytic methods focus on using captured data to directly support instructors and students typically with an immediate feedback loop. Learning analytics faces the problem of improving a learning experience from a more holistic point of view.

Analytics in general refer to a generic set of techniques and algorithms that has been used for quite some time in other domains. When applying analytics to learning, there are certain features that can be directly projected into this new domain, whereas others need some adjustments due mainly to the conditions imposed by the learning experience. Campbell and Oblinger defined analytics in educational contexts as an engine that works in five steps: capture, report, predict, act, and refine (Campbell et al. 2007). The remainder of this document analyzes in detail these five steps, what issues are present in several of them, and their eventual interdependency. Decisions taken at the initial stages may have a profound effect in the following stages. For example, the type of data obtained during the capture stage will likely shape the type of predictions and actions that can be derived. The objective of this paper is to provide a view of the aspects that influence each of these stages and offer a holistic account of the issues that affect the deployment of a learning analytics experience.

The "capture" stage contains the required measures and techniques to make sure the information about events occurring in a learning environment is stored. This information is likely to be contained in a set of heterogeneous sources and is not directly ready to be processed. The issues present in this stage are related to how this information is centralized, how it is encoded, what issues might be encountered when the amount of data is very large, and how to make sure required security measures are observed.

The "report" stage assumes that the data obtained in the previous stage is processed by an arbitrarily complex method ranging from simple visualizations to more complex algorithms that summarize or combine data. The result is new information that is reported back to the stakeholders. Various aspects need to be taken into account in this phase. The frequency of the reports will be affected by the complexity of the processing applied to the data. For example, if the amount of data captured is very large, a real-time computation of the reports is not feasible, whereas a periodic execution is more feasible. Also, the destination of these reports is important. Students, instructors, and administrators are the three groups that have a direct interest on receiving this information. In the case of students, reporting the data connects their activity with self-reflection on the learning experience. For instructors, receiving the reports directly increases their exposure to the intricacies of the learning process fostering the deployment of adjustments. In the case of an administrator, the received reports may also help to understand other issues in a learning community.

The "prediction" stage takes the support for stakeholders further. In this stage the applications are specifically designed to provide answers to previously formulated questions. One common example is the probability of a student failing a course (Romero et al. 2013). These predictions are computed using the data previously collected and applying one of the numerous predicting techniques available. From the point of view of information, predictions can be seen as a more sophisticated report

and as such, it can be distributed as well among the same stakeholders. Prediction algorithms may help students anticipate difficulties during an experience, help instructors to identify students that are not performing as expected, and help administrators to anticipate complications in a course.

The "act" stage is perhaps one of the most sophisticated and relies on the existence of predictions produced in the previous one. The objective is to generate actions that will change any aspect of the learning activity. For example, if the prediction algorithms report that the probability of a student becoming disengaged from a course is high, the material assigned can be automatically adjusted to include more supporting documents that may help alleviate the situation. The target of the acts included in this stage can be any of the stakeholders. A system may decide to send a notification to the counseling services of an institution after detecting that certain students may require their services. In a broader interpretation of this stage, manually deployed actions would also be included. In other words, an instructor, after reviewing the information reported by the learning analytic application, may decide to take some actions to target certain predicted situation. Thus, the actions considered may range from manually to automatically deployed.

The last of the five stages proposed includes the "refinement" of the overall approach. The objective of this stage is to make sure that the previous stages are constantly reviewed and supervised and adjustments are included to improve their suitability. This refinement can be applied to the collection events to improve the quality of the information that is retrieved. The reporting can also be refined by providing more informative information to users. The connection between refinement and prediction algorithms is straightforward as both stages need to reduce the probability of false predictions and increase the accuracy of the results. Finally, the actions that are considered to modify a learning experience can also be the subject of refinement to make sure that they are applied to the right individuals, under the right conditions, and with maximal impact.

The wide variety of learning analytics applications depends on the measures taken in these five categories, with some of them being totally ignored. In its simplest form, a minimalistic learning analytics experience may be envisioned in which data is captured and reported back to the instructors in a learning experience. Although they may take some measures derived from the observations, there is no formal procedure to characterize or document these actions. On the other side of the spectrum, a comprehensive experience could be conceived in which all five stages are taken into account: the stakeholders receive the reports derived from the data, a set of algorithms are in place to produce various predictions in the learning environment, the system creates and deploys certain actions, and there is a regular review of the overall process to improve its effectiveness.

As it can be seen from the two extreme cases previously described, the range of possible learning analytics experiences is very large. This variation is derived not only from the stages that are considered but also by depth of the measures considered at each stage. At the same time, these two dimensions are themselves significantly affected by the type of learning experience being considered. The types of feasible applications are not the same in a face-to-face environment with highly engaged

Table 2.1 Questions to be addressed in each of the learning analytics stage

Stage	Questions
Capture	• What data is being collected? • How frequently is the data collected? • Where is the data going to be stored? • Which format is going to be used to represent all events? • Are the observations *securely* stored?
Report	• Who will receive the reports? • How frequently? • What kind of information needs to be reported? • How will the reports be accessed?
Predict	• Which aspects of the experience need to be predicted? • Which factors can be used as input for the prediction algorithms? • What kind of prediction technique will be used? • How is the accuracy of the prediction going to be measured? • How are the predictions reported to the stakeholders?
Act	• What actions are considered? • How are the actions deployed in the learning environment?
Refine	• Are the data sources appropriate? Are the storage and access requirements for the data appropriate? • Are the produced reports useful? Are they reaching the appropriate stakeholders? • Are the prediction algorithms adequate? Are the predictions useful? Is the accuracy appropriate? • Should the set of actions be revised? Are the actions properly deployed?

students participating in a collaborative experience, or an online experience based on individual tasks, to mention two extremes.

Table 2.1 shows a characterization of the five stages by the type of questions that are asked and answered by each.

The answers to these questions can vary in complexity depending on the aspects considered. For example, when answering the type of actions that are considered, a detailed description of the technological solutions needed for the action to be incorporated to a virtual learning environment may be required. This, in turn, may unfold as the detailed description of the functional requirements needed for such platform to be integrated into an already existing infrastructure.

The questions in Table 2.1 offer a convenient division of the issues to be taken into account when facing the design of a learning experience in which analytics is going to be used and to quickly identify the tradeoffs to consider. A more detailed description of each of the stages follows.

Throughout the rest of the document, a set of specific examples (which in no way pretend to be an exhaustive listing) has been chosen to illustrate the interplay between the five stages.

- Signals (Arnold 2010): Use of data collected in the LMS to predict students in danger of falling behind in a course.
- Macfadyen and Dawson (2010): A predictive model based on statistical regression to detect students at risk based on LMS data.

- GCS (Lin et al. 2009): A text mining application to process forum discussions in an LMS and classify them in different categories.
- CourseVis (Mazza and Dimitrova 2004): Process the information inside the Moodle database regarding interactions of online students and relay that information to instructors.
- Loco-Analyst (Jovanović et al. 2007): A back annotation tool to label the material of a course with the events that were produced in a learning experience.
- Virtual Appliances (Pardo and Delgado Kloos 2011): The use of virtual machines to capture an exhaustive collection of events occurring in a learning environment.
- NetLogo (Blikstein 2011): A tool to monitor students while they write programming assignments and detect complex patterns.
- Detecting Exploratory Dialogs (Ferguson et al. 2013): An application that analyzes the messages written by course participants to identify mental processes.
- iWrite (Calvo et al. 2012): Automatic generation of feedback to writing assignments.
- CAM (Schmitz et al. 2007): A schema to capture and manipulate contextualized attention metadata regarding students participating in a learning experience.
- Tin Can: An application programming interface to collect and store events occurring in learning environments.
- Student Activity Meter (Govaerts et al. 2012): A tool to visualize the events collected in the LMS hosting a learning community.
- Learning Glass (Leony et al. 2012): A tool providing a dashboard with multiple representations of the events previously collected in a learning experience.
- Student Explorer (Lonn et al. 2013): Early warning system (for an engineering mentoring program).
- Sherpa (Bramucci and Gaston 2012): A recommendation engine to send suggestions to students about courses, information, and services.
- Interventions derived from online discussions (Wise and Hausknecht 2013): A system that processes online discussions and derives guidelines that are shared with students and instructors.
- SNAPP (Bakharia and Dawson 2011): A browser plug-in that provides a visualization of the activity in a discussion forum hosted in an LMS.

2.1.1 Data Capture and Representation

The first stage when deploying learning analytics in a learning environment is how to capture and represent the different interactions that occur. This is one of the aspects that influenced the most by the underpinning pedagogical strategy. Due to the influence of technology, a transition is taking place in educational contexts from a content-centric approach to a student-centric one. In the past, the difficulties of accessing the information and documents needed to learn certain skills would require that an instructor spend most of his or her attention preparing and delivering

an educational experience. Instructors had to devote a significant amount of time to find the appropriate material, or create their own, to explain the appropriate concepts. This material was solely found in text books or other written documents, and a physical copy was needed to interact with them. This approach towards teaching and learning has been known as *teacher-centered*.

With the advent of technology, the possibility of accessing the material in various forms, and the high availability offered by the Internet has radically changed the conditions under which experiences are designed and deployed. The focus is no longer somewhere "in" the material, but on how students interact *with* the material, in what is known as *student-centered* education. Student-centered education can often depend upon measures of learning beyond the rote, through what is called as "active learning." Felder and Brent define active learning as "anything course-related that all students in a class session are called upon to do other than simply watching, listening and taking notes" (Felder and Brent 2009). In this context, students are supposed to engage in meaningful activities and think and reflect about what they do. Active learning is a broad term that encompasses various pedagogical strategies such as collaborative learning, project-based learning, and integration with previous activities. All have in common, however, the requirement that the active participation of the students is required for the strategies to prove effective. Previously published studies argue that the effectiveness of these techniques is based on solid empirical observation (Prince 2004). Some authors already include active learning as one of the seven principles to improve undergraduate education (Chickering et al. 1987).

One of the challenges faced when considering the effectiveness of active learning methods is derived from the difficulty of measuring improvements. In order for these methods to succeed, students need to be engaged with the course activities (Fredricks et al. 2004). Student engagement has been identified as one factor that correlates positively with achievement-related outcomes. Fredricks et al. also identified three types of engagement: behavioral, emotional, and cognitive (Fredricks et al. 2004). Engagement is connected to how students feel, think, and behave within a course, and is a good indicator of where apathy might be eliminated and learning enhanced. Additionally, engagement is described as malleable, that is, it can be changed more easily than other individual traits. It is for this reason that observing and measuring the level of engagement of students in a course may offer a powerful indicator to the type of interventions required. A student that barely participates in the activities in a learning environment, does not submit required deliverables, or does not participate at all in discussion forums is more likely to be disengaged. Having a clear and in-advance account of this situation may prompt a modification of the activities, a direct message from the instructor, a suggestion to contact counseling services, a change in the material being covered, etc. But measuring this engagement is riddled with pitfalls. Technology, however, is now offering the possibility of capturing detailed accounts of the events occurring in a learning environment. These events can then be used as the basis to estimate parameters such as level of engagement.

In a study, Anderson and Garrison (1998) proposed an abstract model that captures the type of interactions occurring in a generic learning experience. Students interact with various resources (course notes, multimedia resources, assessments, simulators, etc.), but they interact as well with other students and instructors. In fact, the model considers all possible interactions between these three entities: students, instructors, and material. Later, Anderson claimed that deep and meaningful learning occurs when one of the possible interactions is at its high level. Additionally, if more than one of these interactions is at a high level, the quality of the experience is increased (Anderson 2003; Miyazoe and Anderson 2010). This model isolates the concept of *interactions* as one of the aspects that, when properly captured, can be used as an input into the five stages of learning analytics to increase the effectiveness of a learning experience.

Interest in capturing these observations is not recent. There are pedagogical strategies that already rely on technology to offer detailed information about what is occurring in a learning environment so that instructors can adapt the content accordingly. For example, just-in-time teaching (or simply JiTT) was proposed by Novak (2011) as a strategy to offer a closed loop between the participation of students in activities previous to a lecture, and the material covered in the lecture. Students participate in a set of so-called *warm-up exercises* that are submitted and graded automatically. The results are then forwarded to the instructors to adapt the material or the activities accordingly. This technique can be considered as an example of how the design of an experience can quickly process and disseminate information to a number of varied stakeholders.

The complexity of capturing detailed information about the events occurring in a learning environment is too high to be widely adopted by most individual users or non-specialists. One of the first sources of information about the events occurring while learning is the LMS. These platforms are the communication hub among students and instructors, and are used to share documents, forum posts, and in some cases, blogs, wikis, etc. Some of the actions in these platforms are registered in the internal database. The actions that are recorded and the format in which this data is encoded are highly dependent on the implementation of the platform. But at a basic level, these platforms record every time a user accesses a resource in a course. These resources may include the initial page of a virtual course, the syllabus, course notes, the forum, etc. In recent years there has been a surge in the effort to track student events in different learning scenarios. A comprehensive description of the state of learning analytic is described by Ferguson (2012). One of the early initiatives is the *Signals* tool developed at Purdue University (Arnold 2010) that analyzes large datasets collected through the LMS and predicts which students that are in danger of falling behind in a course. Macfadyen and Dawson analyzed the different factors collected in an LMS and selected the most appropriate to anticipate student performance and created a predictive model to account for the results (Macfadyen and Dawson 2010). Online discussions have also been analyzed by various authors with techniques such as clustering and classification. to infer leading or passive students, the types of messages posted, etc. (Lin et al. 2009). Other more exhaustive approaches have been presented in which detailed feedback is given to the

instructors based on observations collected in different modules of a LMS (assessment, forum, submissions, etc.) (Jovanović et al. 2007). Instructors may browse through course material annotated with information derived from these observations and modify the resources accordingly. Mazza and Dimitriova proposed a system to trace the interactions of "distance learning" students in the Moodle LMS (Mazza and Dimitrova 2004). Their approach analyzes the Moodle database to detect individuals that need particular attention, as well as patterns and access trends. However, the information provided by their system is not specific enough to determine the type of support needed by the students regarding a given topic. Furthermore, their proposal obtains data only from interactions within the LMS and does not take into account resources external to the LMS. As a conclusion, and as stated by many authors, LMSs offer a serious limitation while observing a learning environment (Dawson et al. 2008; Mazza and Dimitrova).

As an example of the type of data that is being used, let us consider the open source LMS Moodle. The administrator of a course using Moodle has the ability to download a *report*. The interface offers various filters depending on parameters such as participants, date, time, and type of activity. The information is encoded as a plain text file. The first line contains the names of the data fields in the file which are:

- *Course*: The name of the Moodle course from which the data has been requested.
- *Time*: The date and time (with minute precision) when the event was recorded.

It should be noted that due to the level of *granularity* here, two identical events may be recorded with identical time while in fact they occurred up to a second apart.

- *IP address*: The IP address of the computer from which the petition was received. This information may be used to locate the user within a certain geographical area.
- *Full name*: The full name of the user as registered in the LMS.
- *Action*: The name of the event recorded. Example of these names are: *discussion mark read*, *forum add discussion*, *forum add post*, *forum delete discussion*, *forum update post*, *forum user report*, *forum view discussion*, *forum view forum*, *forum view forums*, *notes view*, *resource view*, *resource view all*.
- *Information*: Additional information about the event. For example, if the event is *forum view discussion*, this field contains the name of the discussion. In some other cases, for example, the event with name *forum view forums* which is recorded when the user visits the page in which the course forums are listed, the last field is empty.

Each line is encoded as tab-separated values, that is, the values of each field are separated by a tabulator, a format commonly used to transfer values from a database. The file thus contains as many lines as the number of events plus one, the first line that includes the names of the columns.

Similar information can be collected in other LMSs. The differences among LMS platforms are with respect to how the information is reported (plain text, PDF, spreadsheet, etc.), how to select the proper events to include in the report, if the

reports need to be requested by a user interacting with the web interface through a browser, or if there is an Application Programming Interface (API) that can be used by a second application to obtain the data.

Information obtained from an LMS has been used in the past to analyze its correlation with other parameters such as academic performance (see Macfadyen and Dawson 2010 or Tanes et al. 2011 for two representative examples) to deploy what is known as *early warning systems* to detect students that are about to disengage from the course. The extent to which this information is reliable depends substantially on the type of experience. For experiences adopting a distant education approach, the interaction that occurs in the LMS is a significant part of the overall interaction. In face-to-face contexts, or even blended learning, the LMS may capture only a fraction of this interaction, thus affecting the reliability of the predictions that are derived from the data.

But the underlying technique of using technology to mediate interactions and record the events that occur can be taken beyond the LMS. A more generic scenario can be envisioned in which interactions are mediated by a set of tools previously prepared to record events. There have been some initial research initiatives in this direction inspired by tracking events in other contexts. The idea is to collect events from an heterogeneous set of sources and provide a common representation so that the data can be further processed. These sources may include specific tools such as an interactive development environment, social platforms (Twitter), or an entire set of tools packaged as part of a virtual appliance.

For example, Wolpers et al. presented a framework (Wolpers et al. 2007) to capture what users are paying attention to while working on a personal computer. The approach required instrumenting the device so that every time a user executes a tool part of the Office suite, opens a page with the web browser, plays a multimedia file, or uses a predefined set of communication applications, the events are recorded. In their work, the authors acknowledge the potential of analyzing these events and detect patterns as to what the user is paying attention to. They acknowledge the challenge of capturing the information from such a large set of tools while representing the events in a common format. To solve this problem, the authors propose the use of the *Contextualized Attention Metadata* (CAM) format (Schmitz et al. 2007), an XML-based representation capturing the required information.

We presented a similar approach based on the use of virtual appliances in the context of learning scenarios to restrict the tools that are being observed and provide a self-contained and clearly delimited environment for students (Pardo and Delgado Kloos 2011; Romero Zaldı́var et al. 2012). With this technique, a virtual appliance is installed by students on their personal computer. The appliance is fully configured with the tools required to participate in the course activities, tools for exchanging documents using a version control system, and other conventional tools such as a browser and an editor. All of them are instrumented so that each time they are used, an event is recorded. The collection of events is then relayed to a central server as part of the transactions originating from the version control system. Figure 2.1 illustrates this approach.

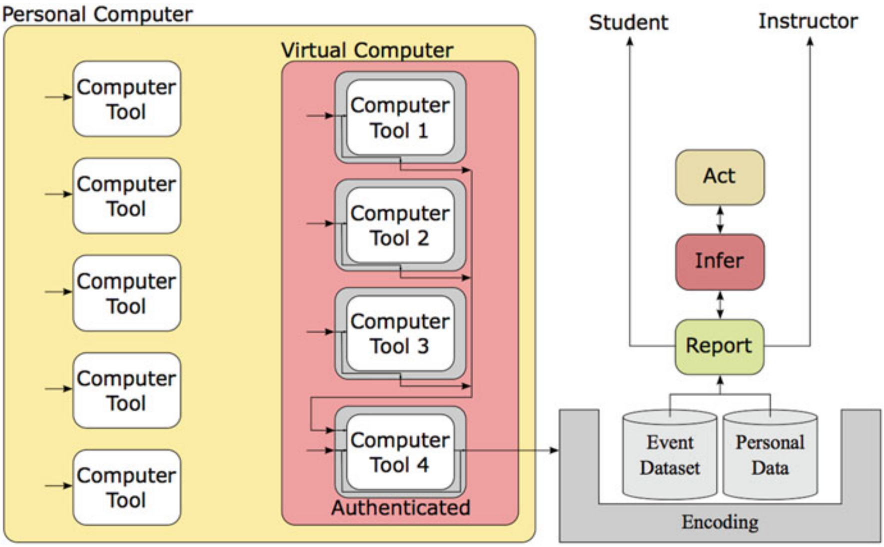

Fig. 2.1 Instrumentation of tools in a virtual appliance

The fact that the tools are contained in a virtual appliance that needs to be started and stopped as any other regular application offers the advantage that the recorded events exclusively target the actions taken while the students are performing work for a course, as opposed to more generic approaches that track every event occurring on a computer.

Instrumentation of specific tools has also been used in other contexts in which a significant amount of work is performed with an application. For example, Blikstein proposes the use of NetLogo to observe students while they write programming assignments (Blikstein 2011). The platform captures key strokes, button clicks, and changes in the files being edited. This level of detail produces large quantities of events that need to be represented, stored, and manipulated with additional tools and techniques. Additionally, the detailed information also serves to detect patterns that are difficult to generalize, as they are tightly coupled with the structure and topic of the activities under observation.

Basic text written by users is also an increasingly popular source of information used for analysis. For example, Ferguson et al. used the text obtained from a synchronous online discussion to detect patterns suggesting exploratory dialog among the participants in a distant course (Ferguson and Shum 2011; Ferguson et al. 2013). Dialogs between participants in a course are considered more revealing of cognitive activity than monologic material. Focusing on specifically revealing cognitive tasks was also the objective proposed by Calvo et al. (2012). After students are given a writing assignment, a tool used for writing the document is instrumented to track the events that occur during the activity. The final text is analyzed and an automatically generated document containing feedback is sent back to the user. The events are also analyzed to detect different levels of engagement in task.

These last two examples can be characterized as providing deep and detailed accounts of a very specific set of events occurring in a unique activity (writing a document, online synchronous discussion). On the opposite side of the spectrum are other applications that use a variety of tools and platforms to collect evidence of a user's activity. For example, Santos et al. propose monitoring how students self-reflect in their activities based in a combination of self-reported data, and comments collected from social networks such as Twitter (Santos et al. 2013). By triangulating data from multiple sources, it is likely that the quality of the predictions is increased.

2.1.1.1 Representing the Captured Events

In the examples previously described in which events are collected from a wide variety of sources, there is an important issue that deserves special attention: the representation of the captured events. When a single source of events is used, the problem is not important and can be easily solved with a notation specifically designed to capture the events. However, when the number of sources is larger, or even in scenarios in which events can appear from unexpected sources in the middle of an experience, a unified representation of events is needed.

Several research initiatives have appeared in an effort to resolve this issue, and it is interesting to analyze the tendency of the solutions. The first attempt to solve this problem was proposed by Wolpers et al. in the more general context of analyzing attention in user applications (Wolpers et al. 2007) as previously discussed. The initial version of the *Contextual Attention Metadata* (CAM) resulted, and its initial schema contained around 30 entities that could be applied to capture the information attached to an event. This information was divided in other entities such as items, actions, and sessions.

CAM was then used in other contexts such as self-regulated learning, personalization, or self-reflection (Scheffel et al. 2010; Govaerts et al. 2012; Muñoz-Merino et al. 2010) but these uses all required simplifications of the initial notation. The key observation that was derived from this evolution is that in order for a data representation format to be highly generic, it needed to reduce its complexity to the bare minimum while retaining its expressive power. The expressive power was needed to be able to apply more sophisticated querying mechanisms such as those described in (Muñoz-Merino et al. 2011).

The ADL-sponsored project Tin Can has pushed this evolution to its next stage. The project starts by acknowledging that activities occur in various contexts and are mediated by a variety of tools. The focus is shifted from organizing a collection of resources to collecting the events derived from the interactions emerging among *all* the entities in the learning experience.

The project can be seen to be addressing the problem of how to capture these events with a generic representation. The proposal cannot be any simpler. Every event is captured by a triple of *Actor*, *Verb*, and *Object*. With this representation, events become sentences in which a certain fact is explained. The simplicity of the formalism allows for a wide variety of events to be encoded with these triplets. To

encode more complex events, Tin Can breaks them down into simpler actions that are then encoded in the proposed syntax. The expressiveness of this approach relies on the capacity for generic agents to process large numbers of basic events and *deduct* more complex patterns.

The formulation has also resulted in a change of terminology when designating different agents in a learning environment. Aside from the LMS, a new entity appears in charge of collecting these triplets, the *Learning Record Store* or LRS. Additionally, the proposed scheme does not assume that users interact with resources through a browser. In fact, the *actor* of a triplet does not need to be a person, but it can be an application. Thus, when two generic entities (students, instructors, applications, or content) interact through an application, a triplet can be relayed and stored in the designated LRS.

Although the project is in its infancy, it remains to be seen the impact it will have on how exhaustively events are collected from learning environments and how applications can make use of the notation to derive insights and interventions. The simplicity of the approach also delegates responsibility for semantic arrangements to the agents processing the events. The vocabulary and meaning of the triplets can be defined by each application separately. Thus, two experiences taking place in two radically different scenarios with different content organization, pedagogical strategy, and terminology can still encode their observations in this scheme. The applications processing these events, though, must be aware of all these aspects to produce meaningful insights. With this bottom-up approach, the adoption barrier of this paradigm is basically removed, and the complexity shifted to higher order agents, thus fostering the potential for sharing and exchanging information across experiences. Delegations of responsibility like this one are a crucial consideration for the subsequent stage of learning analytics, reporting.

2.1.2 Report

Reporting is the second stage of the learning analytics engine. By reporting we refer to relaying the raw data obtained from a learning environment or the procedures that manipulate this data back to the stakeholders. An important portion of the reports relies on visual methods to convey the information. The reason for this bias is due to the potentially large amount of data collected. If only a small subset of students is observed, a direct presentation of the data might be feasible. But when a large number of events are collected on a large population of students, the information needs to be processed to create a more intuitive representation that can be understood in a few seconds.

The area of information visualization has been the center of numerous research activities (see Kerren et al. 2008 for a survey). The objective is to cope with the so-called *information overload* problem in which users receive too much information that precludes them from interpreting it. Visualizing events occurring while students participate in a learning experience is very close to current applications that offer

the possibility of tracking your physical activity, computer use, etc. As pointed out by Duval (2011), these applications offer a rich context from which to derive techniques and ideas to apply to the area of learning.

Visualizations can have variable scopes. Dashboards are used to combine multiple graphic elements that each focus on a single aspect.

There are numerous examples of systems that monitor student involvement in a distance education course (see Mazza and Milani 2004 for an example). These tools obtain the data from the web log data generated by a LMS and provide information for instructors identifying and preventing problems associated with distance learning. Santos et al. proposed another application that would process events stored in CAM and produce a dashboard with graphical information about the activities of users in a PLE (Santos et al. 2011). Although the ultimate goal was to improve the functionality of this environment, due to the similarity with conventional interventions in learning experiences, it can be considered a learning analytics visualization platform. A more refined application targeting a learning environment was proposed by the same authors to specifically visualize goal achievement (Santos et al. 2012).

We also implemented another example of dashboard visualization specifically designed for learning environments (Leony et al. 2012). The application offers the possibility of combining various graphical representations encapsulated as widgets in a canvas. Each of these widgets can then be configured and arranged by users independently according to their needs. Figure 2.2 shows an example of a dashboard with four visualizations. The tool also allows for events to be filtered by student, time, or even groups of students.

Another example of a dashboard relaying information about a learning experience was proposed by Govaerts et al. (2012) (in which the author participated). The Student Activity Meter (SAM) assists students and instructors to increase self-awareness of their participation in a course. Figure 2.3 shows an example of a SAM dashboard for a course. The information is represented in three main areas. In the upper left corner, the overall number of events is shown as accumulated activity for each of the students. The bottom left area allows browsing through the course documents sorted by decreasing number of accesses. This information allows a quick detection of those documents that are being accessed the most. The upper right corner shows parallel coordinates that allow a side-by-side comparison of multiple factors to identify trends. The bottom right visualization offers the possibility of grouping various types of events into bins and rendering them as histograms.

Another visualization that has been successfully deployed for a large number of students in a higher educational institution is part of the Signals system (Arnold 2010). The system collects information from instructional tools (like a LMS) to try to detect students at risk of abandoning a course. As part of the actions derived from the data, students are shown a traffic light with a color depending on their estimated status with respect to the course. Instructors are also shown a listing of the status for all the students in the course. These visualizations, although much simpler than a dashboard, distill all the data and produce a single entity to remove the task of interpreting data from the end users.

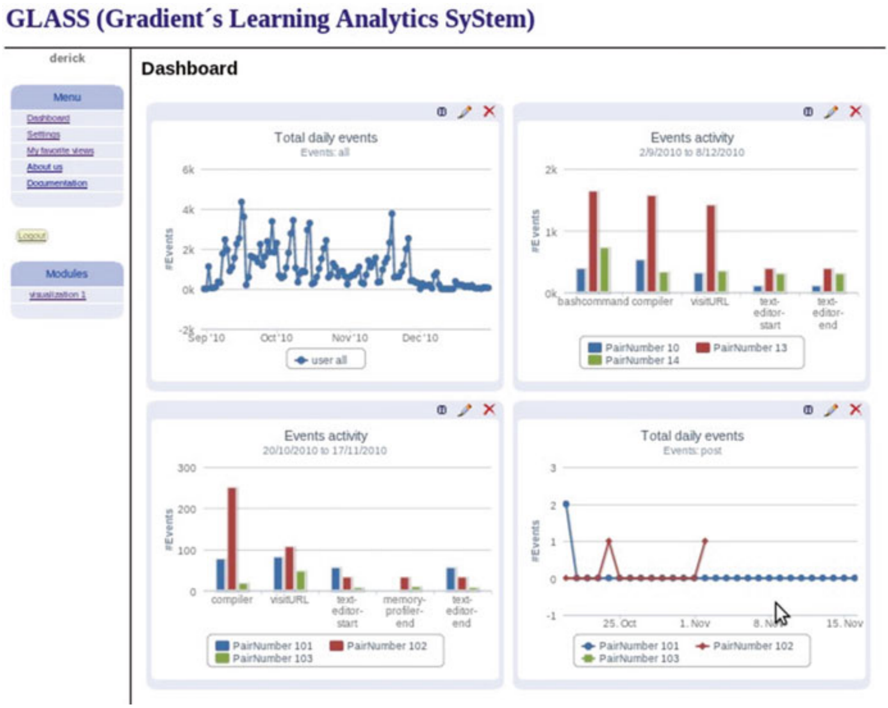

Fig. 2.2 Our dashboard with four widgets to visualize different types of events (Leony et al. 2012)

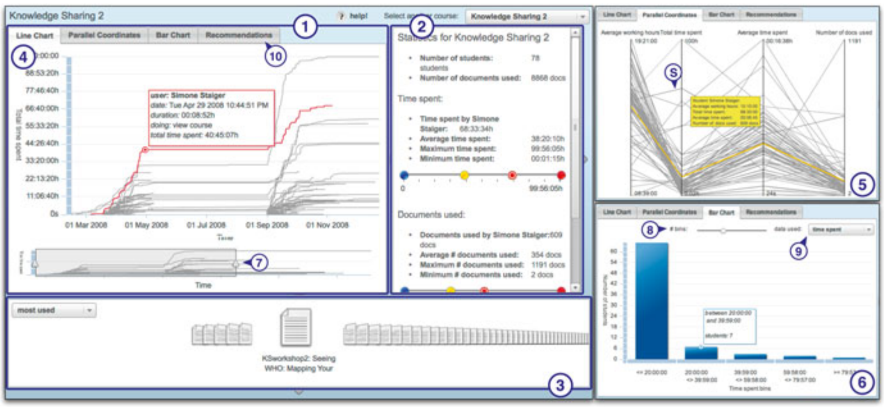

Fig. 2.3 Dashboard of the student activity meter as proposed in Govaerts et al. (2012)

Other examples of dashboard visualizations have been recently proposed (Essa and Ayad 2012; Kump et al. 2012; Rivera-Pelayo et al. 2013). But dashboards represent a category of visualizations in which the emphasis is on achieving a comprehensive view of a large number of events. There are other visualization strategies

that provide a detailed view of a single aspect in a learning experience. For example, SNAPP is an interaction diagnostic tool that analyzes the participation of students over time in discussion forums (Bakharia and Dawson 2011). The information is visualized embedded in a page of the LMS by using a client-side browser extension. This solution relies on users to install the extension, requiring no direct modification of the LMS. As a consequence, the extension requires multiple versions for the different browsers.

The challenge when designing and deploying visualizations is to accurately measure their true impact. Typical strategies to validate this technology rely on surveys of students because of how difficult is to detect variations in behavior between users using the visualization and users ignoring it.

2.1.3 Predict

The next step in the learning analytics engine is prediction. After data has been captured regarding the events occurring in a learning environment, the true power of this data is to derive models capable of anticipating events that will occur in the future. There are numerous techniques and methods that can be used to create these predictive models, and choosing the most appropriate method depends on the environment in which they are applied. A large percentage of the possible solutions for prediction is covered by two categories: statistical inference and machine learning. In both cases the objective is to use the collected evidence to derive a model to predict future events.

As illustrated in Fig. 2.4, the data collected can be generically grouped in a set of factors $\{f_1, \ldots, f_{k+1}\}$. One or several of these factors are selected for prediction. For the sake of simplicity, let f_{k+1} be the single factor to predict. The remaining factors $\{f_1, \ldots, f_k\}$ are combined to obtain the prediction model. Once obtained, factor f_{k+1} is then used to validate the accuracy of the model.

A comprehensive description of statistical inference methods is beyond the scope of this document. Focusing on the context of learning analytics experiences, there are several aspects that need to be taken into account. The data derived from learning environments may have different types of statistical distributions. The appropriate inference method to apply depends on the type of probability distribution in the collected factors. As a consequence, a preliminary analysis is required to determine these distributions. Although learning analytics platforms may capture a wide variety of event types, a significant amount has discrete values. Assessment scores, number of accesses to resources, posts in a discussion forum, etc. are all discrete variables with a reduced number of possible values in contrast with continuous variables that appear in other areas (such as, for instance, temperature and pressure)

One of the simplest prediction models is linear regression. The objective is to derive a linear combination of the factors $\{f_1, \ldots, f_k\}$ that estimates the value of the factor f_{k+1}. The line derived from the linear combination of the factors minimizes the sum of the squares of the differences of all the data points. If the factor f_{k+1} is

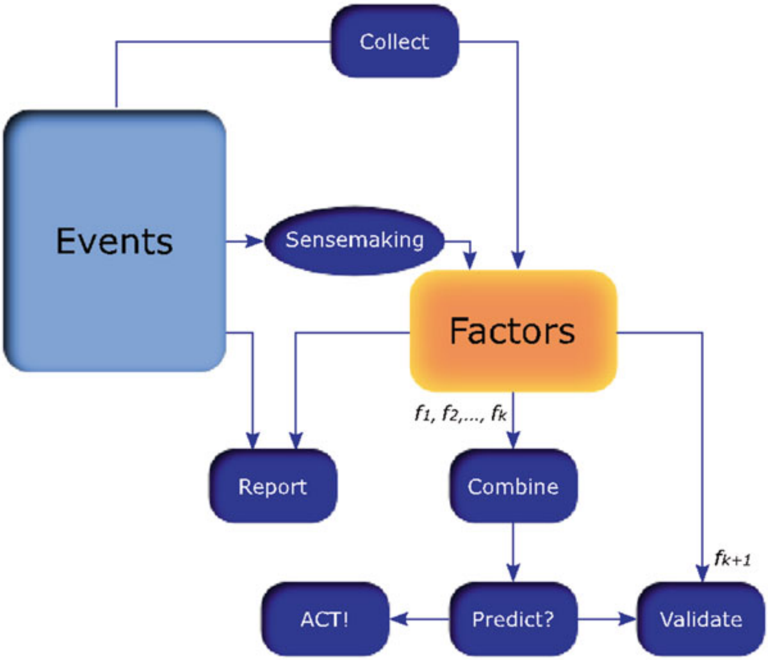

Fig. 2.4 Steps of learning analytics

categorical, then a variant known as *logistic regression* is used. The process to derive a linear model can be summarized in three steps: verify that the factors used for the regression have the correct statistical distributions, calculate the linear expression, and verify the significance of the obtained expression.

A second set of techniques used for prediction is what is known as machine learning. In general, a set of factors is used to create an algorithm that is then applied to additional data to predict one or several of these factors. The fact that the final product of a machine learning algorithm is in itself another algorithm (although one with a very specific purpose) has led to the proliferation of myriad alternatives. After choosing the appropriate representation for both the input data and the derived algorithm, a function is needed to calculate how effective the produced algorithm proves. This function is essential to guide the search towards the solutions that offer the best predictions. The third ingredient required is a method to create alternatives to search for the final solution. These three ingredients have multiple possibilities translating into an even large number of combinations all of them considered *machine learning techniques* (Domingos 2012).

The data collected is used to *train* the algorithm to predict one of its factors. Once the algorithm is produced, it can be applied to a new set of data and the prediction of the selected factor is the produced result. The main problem of these techniques is known as over-fitting, that is, trying to capture the structure of a set of events with the wrong model. In this case, the resulting model ends up not

predicting the result accurately when used in new data. In order to avoid this pitfall, additional techniques such as cross-validation are applied to the resulting model to gauge the need for additional training data.

Aside from these two categories, other learning scenarios may require ad-hoc techniques for prediction. For example, when detecting if exploratory dialog occurs in a learning activity, a specific set of words is detected automatically, and then a manual processing stage is performed to identify the indicators (Ferguson and Shum 2011). Given the wide variety of activities in a learning environment as well as the various types of data that can be captured, ad-hoc techniques may provide a more adequate alternative than statistical regression or machine learning.

The design of an experience in which learning analytics is used needs to take into account the possibilities offered by technology to observe and collect detailed accounts of the interactions occurring between the different stakeholders. Some initial explorations of how to embed Learning Analytics in educational design have been published (Lockyer et al. 2013; Jovanović et al. 2008). But the area still lacks comprehensive studies in which the information collected is then properly related to the activities and resources within a course. Taking students and teachers as the main stakeholders forces the appearance of new issues. What kind of events can be observed when the information is reported to instructors? Can the material be adapted to be able to monitor aspects that facilitate fast instructor feedback? Could pedagogical aspects such as the presence of misconceptions be detected and tackled? What is the best way to report the collected information to students? Should results from a student be compared with other students, or the rest of the class? How can the privacy and security of the stakeholders be properly managed? All these questions are inherent to the holistic focus of learning analytics. In fact, a successful use of learning analytics techniques must address all these factors. A tangible improvement on the overall learning objectives is the obvious goal, but other aspects such as its feasibility in terms of deployment, its adequacy to the current pedagogical structure, or its compliance with privacy issues must also be taken into account to accurately assess the benefit of an experience.

2.1.4 Act

If prediction in learning analytics can be done with a wide range of techniques, the next stage, acting, offers an even a richer set of alternatives. By *acting* we refer to the manual or automatic actions in the learning scenario that are derived from processing the collected data. These actions may range from something as informal as talking to a student to a significant change automatically introduced in the course material. Thus, one variable to take into account is the involvement of instructors in these actions. Concrete solutions may rely completely on the instructors, may be totally automated, or may be any hybrid solution in between.

Learning analytics is different from other research areas such as intelligent tutoring systems in the sense that the actions derived in this stage tend to be more

independent of the previous information gathering scenario. For example, the Signals project used the collected data to then provide students with generic feedback. In other words, rather than design actions that depend on the topic of the course, a generic action of providing feedback to students was analyzed (Tanes et al. 2011). The intrinsic effect of this technique has a low dependence on the type of learning experience. The action of providing feedback to students depending on the predictions of their performance can be generalized to a wide number of scenarios.

A similar philosophy of maintaining the teaching staff as part of the overall approach was adopted by the early warning system *Student Explorer* (Lonn et al. 2012, 2013). The system displays whether instructors should encourage students, explore their progress in more detail, or engage with students to assess possible academic difficulties. The tool has been deployed following a gradual approach in which the first edition was made available to a limited number of instructors with a simple integration procedure with the institutional LMS and other corporate applications. Once the tool was validated, a second stage was carried out to scale it to a large number of users. This strategy has the advantage of detecting bottlenecks and issues early in the design flow before the application is used across an entire institution.

Other tools and initiatives specialize in sending messages to students in a wide variety of formats. For example, Sherpa is an application that collects as much information about students in the South Orange County Community College District and executes actions to help students when selecting courses, become aware of events, or to help extend support to at-risk students (Bramucci and Gaston 2012). The actions are generically named *nudges*. The nudges are programmed into the system by combining factors with a set of Boolean and numeric operators. These rules are constantly evaluated to detect triggering situations that result in a message sent to the user regarding available open courses, generic college information, and personalized links to certain services.

Wise et al. proposed a novel intervention in which the result of the analysis stage is shared with the students as part of yet another learning activity within the experience (Wise and Hausknecht 2013). The proposed system provides students and instructors with a visualization of the activities derived from a discussion forum. This is an example of how to use the result as yet another course resource for students to reflect on their learning and to decide their own actions. The difficulty in this case is in closing the feedback loop by portraying the process of analyzing the results and taking actions derived on these results as a pedagogically meaningful activity towards achieving the course objectives. The proposed solution adopted a holistic approach in which the overall design of the activity assumed the presence of a reporting entity capable of making students aware of their participation. As a consequence, the goals of the activity included aspects related to reading posts from the classmates, commenting on them, etc. The self-reflection allowed students not only to validate already existing perceptions with empirical data but also to discover the distance between their subjective perception and the objective measures.

The examples previously described are simply a small sample of the type of actions that may be derived from the analysis of events and highlight the large

number of possibilities to intervene in a learning experience. Although there is growing interest in the area of learning analytics in general, the number of research initiatives that propose new techniques in this stage is significantly lower when compared with collection or analysis technique. This fact shows how the number of barriers to overcome increases when trying to close the feedback loop with the students.

2.1.5 Refine

The last stage of learning analytics design and implementation is more a philosophical proposition than a set of concrete techniques to apply. The main observation at this stage is that once a learning experience is using analytics to influence its deployment, this influence needs to be closely observed for sustained improvements. Collected data, analysis techniques, predictions, and actions are all affected by a high number of external factors that are likely to change over time. Estimating the state of a student when participating in a experience is a very complex and delicate task and as such it is subject to potential pitfalls. In parallel with the use of these techniques there must be a second task of evaluating the framework itself.

Techniques discussed in Sect. 2.1.3 such as machine learning are heavily influenced by the testing dataset, that is, the set of observations extracted from the experience and used to deduce the algorithm used for prediction. If the conditions in the environment change, the model is no longer valid and the predictions will not be reliable. Similar reasoning can apply for predictions based on linear models created with regression techniques. The quality of the prediction is directly affected by the similarity between the newly collected data and what was used for the creation of the model.

Learning analytics is still in its infancy, and therefore, the work in this stage has yet to have a solid presence in the latest research efforts. Once the technology consolidates and the number of applications in educational institutions is large enough, the refinement stage will be studied and explored in detail to propose new methodologies.

2.1.6 Ethical and Privacy Issues

Ethical and privacy issues can be seen as orthogonal to the five steps previously described. The rapid increase of the type of events that can be recorded and the slower pace at which rules and regulations are proposed, has created some tension around how to address these aspects of analytics. On one hand, data collected about personal traits such as gender, personal income, and location clearly belong to the private realm. The same is argued with respect to scores and assessment results. The emerging trend in laws regulating privacy is that users should own the data collected about them, and

institutions are merely *borrowing* this data temporarily for a clearly stated purpose. However, this spirit can be clearly applied to the collected data, but its boundaries dilute when data is processed or shared among several stakeholders. For example, a significant part of learning analytics techniques relies on combining data from a population of subjects to produce a prediction model. Who is the owner of such model? Users have provided the raw material from which it was derived, but the process to create such model was carried out by a third party. Who is the owner of such a data model when an institution gives the collected data to a third party for that purpose?

Legislation is being proposed in several countries to delimit the collection and usage of private data in the context of the Internet (see The White House 2012; Australian Government 2006; EUP 2002; Canadian Standards Association 2001). But these regulations apply to the generic context of users interacting with information and communication technology. When considering students participating in a learning experience, these regulations only provide a generic framework, and do not address numerous dilemmas. Nowadays, the proposals to address these issues depend on multiple implicit assumptions as acknowledged by Slade and Prinsloo (2013). In their work, they propose six principles to be taken into account by higher education institutions to approach these issues: learning analytics is a moral practice and should focus on understanding rather than measuring, students are central agents and collaborators, student identity and performance are dynamic magnitudes, student success is a complex and multidimensional phenomenon, data should be collected and processed with total transparency, and higher education cannot afford not to use data.

Ethics and privacy can be approached at different levels of distance, and researchers must provide multiple visions at all these levels to charter a solid connection between generic principles and concrete suggestions observed by designers, developers, instructors, and administrators. A good example of this connection can be obtained by exploring how these issues have been solved in other areas such as business analytics (Schwartz 2011). A significant part of the issues already addressed in similar areas will likely apply to a learning experience using analytics.

2.2 Conclusions

Learning analytics are now being used effectively in an increasing number of scenarios. The division of analytics into five stages has been used to see the variety of aspects that need to be taken into account, as well as the interdependency between these stages. A more detailed collection of the events in a learning experience is more challenging, but will likely offer a better background for data analysis and prediction. The predicting mechanism needs to rely on a clear channel to relay the information back to students and/or instructors. In each of the stages, a selected set of applications were described to illustrate the variety of options being explored, as well as to bring the richness of solutions in each area to the forefront.

The last stage, refinement, is perhaps the one least documented in the literature. The area of learning analytics has not seen a definitive advance on how all aspects

of analytics processes are refined. A systematic approach is missing, one that illustrates how an effective learning analytics paradigm, one deployed, can be evaluated with respect to a set of indicators, and what comparisons prompt adjustments and re-assessment of the other stages. The interdependency between the phases of learning analytic techniques and learning design is one of the most promising aspects to which refinement can be deployed. The influence of learning activities on the type of data and predictions that can be derived is trivial. However, it is now possible to consider the relationship in the opposite direction. Having the possibility to collect, analyze, and predict certain aspects could potentially influence how the activities are conceived and designed.

Ethical and privacy considerations are another area in which the refinement stage needs to be enhanced. In a society in which privacy is being re-defined, learning analytic experiences must adapt to this situation by embedding a continuous assessment about ethical and privacy issues to better suit the needs of the main stakeholders. A full development of this stage will be the unequivocal sign that learning analytics has finally become an inextricable part of *any* learning experience.

References

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction equivalency of interaction. *The International Review of Research in Open and Distance Learning, 4*(2), 1–14.

Anderson, T. D., & Garrison, D. R. (1998). Learning in a networked world: New roles and responsibilities. In C. Gibson (Ed.), *Distance learners in higher education* (Chap. 6, pp. 97–112). Madison, WI: Atwood Publishing.

Arnold, K. (2010). Signals: Applying academic analytics. *EDUCAUSE Quarterly, 33*(1), 10.

Australian Government. (2006). *Private sector information sheet 1A—National privacy principles*. Sydney, Australia: Office of the Privacy Commissioner.

Bakharia, A., & Dawson, S. (2011). SNAPP: A bird's-eye view of temporal participant interaction. In *Proceedings of the First International Conference on Learning Analytics and Knowledge—LAK'11* (pp. 168–173). New York, NY: ACM Press.

Blikstein, P. (2011). Using learning analytics to assess students behavior in open ended programming tasks. In *Proceedings of the First International Conference on Learning Analytics and Knowledge*, Banff, Canada.

Bramucci, R., & Gaston, J. (2012). Sherpa: Increasing student success with a recommendation engine. In *International Conference on Learning Analytics and Knowledge* (pp. 82–83). New York, NY: ACM Press.

Calvo, R., Aditomo, A., Southavilay, V., & Yacef, K. (2012). The use of text and process mining techniques to study the impact of feedback on students' writing processes. In *International Conference of the Learning Sciences* (pp. 416–423), Sydney.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics. *Educause Review, 42*, 1–24. EDUCAUSE White Paper.

Canadian Standards Association. (2001). *Model code for the protection of personal information. Technical report*. Mississauga: Canadian Standards Association.

Chickering, A., Gamson, Z., & Poulsen, S. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin, 39*(7), 3–7.

Dawson, S., McWilliam, E., & Tan, J. (2008). Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. In *Proceedings of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 231–230), Auckland, NZ.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78–87.

Duval, E. (2011). Attention please!: Learning analytics for visualization and recommendation. In *International Conference on Learning Analytics* (pp. 9–12). New York, NY: ACM Press.

Essa, A., & Ayad, H. (2012). Student success system: Risk analytics and data visualization using ensembles of predictive models. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge—LAK'12* (pp. 158–161). New York, NY: ACM Press.

EUP. (2002). *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector*.

Felder, R. M., & Brent, R. (2009). Active learning: An introduction. *ASQ Higher Education Brief, 2*(4).

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges. Technical report*. Milton Keynes, UK: Knowledge Media Institute.

Ferguson, R., & Shum, S. B. (2011). Learning analytics to identify exploratory dialogue within synchronous text chat. In *Proceedings of the First International Conference on Learning Analytics and Knowledge—LAK'11* (p. 99). New York, NY: ACM Press.

Ferguson, R., Wei, Z., He, Y., & Buckingham Shum, S. (2013). An evaluation of learning analytics to identify exploratory dialogue in online discussions. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge—LAK'13* (p. 85). New York, NY: ACM Press.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. In: *ACM SIGCHI International Conference on Human Factors in Computing Systems* (pp. 869–884), Austin.

Jovanović, J., Gašević, D., Brooks, C., Devedžić, V., Hatala, M., Eap, T., & Richards, G. (2007). Using Semantic Web technologies to analyze learning content. *IEEE Internet Computing, 11*(5), 45–53.

Jovanović, J., Gašević, D., Brooks, C., Devedzic, V., Hatala, M., Eap, T., & Richards, G. (2008). LOCO-analyst: Semantic Web technologies in learning content usage analysis. *International Journal of Continuing Engineering Education and Lifelong Learning, 18*(1), 54.

Kerren, A., Stasko, J. T., Fekete, J. D., & North, C. (Eds.). (2008). *Information visualization: Human-centered issues and perspectives. LNCS Vol 4950*. Heidelberg: Springer.

Kump, B., Seifert, C., Beham, G., Lindstaedt, S. N., & Ley, T. (2012). Seeing what the system thinks you know. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge—LAK'12* (p. 153). New York, NY: ACM Press.

Leony, D., Pardo, A., de La Fuente Valentin, L., Sánchez de Castro, D., & Delgado Kloos, C. (2012). GLASS: A learning analytics visualization tool. In *International Conference on Learning Analytics and Knowledge* (pp. 162–163). New York, NY: ACM.

Lin, F., Hsieh, L., & Chuang, F. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education, 52*(2), 481–495.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439–1459.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 48*(5), 31–40.

Lonn, S., Aguilar, S., & Teasley, S. (2013). Issues, challenges, and lessons learned when scaling up a learning analytics intervention. In *International Conference on Learning Analytics and Knowledge* (pp. 8–12). New York, NY: ACM Press.

Lonn, S., Krumm, A. E., Waddington, R. J., & Teasley, S.D. (2012). Bridging the gap from knowledge to action: Putting analytics in the hands of academic advisors. In *International Conference on Learning Analytics and Knowledge* (pp. 184–182). New York, NY: ACM Press.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education, 54*(2), 588–599.

Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies, 65*(2), 125–139.

Mazza, R., & Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. In *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters* (pp. 154–161). New York, NY: ACM.

Mazza, R., & Milani, C. (2004). Gismo: A graphical interactive student monitoring tool for course management systems. In *TEL'04 Technology Enhanced Learning'04*, Milan.

Miyazoe, T., & Anderson, T. D. (2010). The interaction equivalency theorem. *Journal of Interactive Online Learning, 9*(2), 94–104.

Muñoz-Merino, P. J., Pardo, A., Scheffel, M., Niemann, K., Wolpers, M., Leony, D., et al. (2011). An ontological framework for adaptive feedback to support students while programming. In *Proceedings of the 10th International Semantic Web Conference (ISWC)*, Bonn, Germany.

Muñoz-Merino, P. J., Delgado Kloos, C., Muñoz-Organero, M., Wolpers, M., & Friedrich, M. (2010). An approach for the personalization of exercises based on Contextualized Attention Metadata and Semantic Web technologies. In *IEEE International Conference on Advanced Learning Technologies*, Athens, Greece.

Novak, G. M. (2011). Just-in-time teaching. *New Directions for Teaching and Learning, 2011*(128), 63–73.

Pardo, A., & Delgado Kloos, C. (2011). Stepping out of the box. Towards analytics outside the learning management system. In *International Conference on Learning Analytics* (pp. 163–162). New York, NY: ACM.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(July), 223–232.

Rivera-Pelayo, V., Munk, J., Zacharias, V., & Braun, S. (2013). Live interest meter. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge—LAK'13* (p. 23). New York, NY: ACM Press.

Romero Zaldı́var, V. A., Pardo, A., Burgos, D., & Delgado Kloos, C. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education, 58*(4), 1058–1067.

Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education, 68*, 458–472. Doi:10.1016/j.compedu.2013.06.009. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0360131513001607

Santos, J. L., Govaerts, S., Verbert, K., & Duval, E. (2012). Goal-oriented visualizations of activity tracking. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge—LAK'12* (p. 143). New York, NY: ACM Press.

Santos, J., Verbert, K., Govaerts, S., & Duval, E. (2011). Visualizing PLE usage. In *Proceedings of EFE-PLE11*, La Clusaz, France.

Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2013). Addressing learner issues with StepUp! In *Proceedings of the Third International Conference on Learning Analytics and Knowledge—LAK'13* (p. 14). New York, NY: ACM Press.

Scheffel, M., Beer, F., & Wolpers, M. (2010). Analysing contextualized attention metadata for self-regulated learning. In *International Conference on Computer Supported Education* (pp. 341–346), Valencia, Spain.

Schmitz, H. C., Wolpers, M., Kirschenmann, U., & Niemann, K. (2007). Contextualized attention metadata. In: *Human attention in digital environments* (Vol. 13, Chap. 8). Cambridge: Cambridge University Press.

Schwartz, P. M. (2011). Privacy, ethics, and analytics. *IEEE Security & Privacy, 9*, 66–69.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*. doi:10.1177/0002764213479366.

Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using signals for appropriate feedback: Perceptions and practices. *Computers & Education, 57*(4), 2414–2422.
The White House (2012). *Consumer data privacy in a networked world. Technical report*. Washington, DC: The White House.
Wise, A. F., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In *International Conference on Learning Analytics and Knowledge* (pp. 48–56). New York, NY: ACM Press.
Wolpers, M., Najjar, J., Verbert, K., & Duval, E. (2007). Tracking actual usage: The attention metadata approach. *Journal of Technology Education & Society, 10*(3), 106–121.

Chapter 3
Harnessing the Currents of the Digital Ocean

John T. Behrens and Kristen E. DiCerbo

3.1 Introduction

Recently, DiCerbo and Behrens (2012) suggested the term "Digital Ocean" to describe the emerging reality of ubiquitous and unobtrusive data generated from the use of digital devices in daily life, a notion they contrast against the pre-digital world of expensive and relatively rare data which they characterize as the "digital desert." While originally formulated in the context of the impact of these shifts on assessment argument and use, we extend the discussion to the broader context of data-based educational research and learning analytics in general. This is accomplished in five sections, each of which touches on a shift in perspective or activity that is part of the change as we understand it. In the first section, the experiential aspects of the shift are discussed following DiCerbo and Behrens (2012) in terms of the rise of sensors in digital life. Next, conceptual shifts in understanding educational assessment and educational research data are suggested to provide appropriate conceptual tools for the new and emerging realities. The third section discusses the impacts of the privileged properties of computing for learning analytics. A fourth section discusses issues related to the organization and conduct of research given these shift and addresses implications for the training of educational researchers. Caveats and cautions are presented before concluding remarks.

Following the analogy of DiCerbo and Behrens (2012), we consider ourselves to be on the digital shore: a place in the history of human cultural evolution between the digital desert of the past and the digital ocean of the future. From this epistemic position, discussion of the near past may seem a simple caricature and discussion of the future mere fantasy. However, because the revolution in computing in which we are embedded concerns the transformation of information from physical form and activity to a liquid digital form that can be moved, transformed, synthesized,

J.T. Behrens (✉) • K.E. DiCerbo
Pearson, Notre Dame, IN 46556, USA
e-mail: john.behrens@pearson.com

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_3, © Springer Science+Business Media New York 2014

and acted upon by automated systems (Mislevy et al. 2012), it is also a revolution in the nature of human intellectual and cultural history. It will be, we believe, a fundamental lens through which activity will be understood in the next 100 years, in the same way questioning the role of the individual and the value of systematic inquiry was a central lens in the age of the Enlightenment.

3.2 Experiential Shift: Sensors

The starting point for the conversation regarding the shift from digital desert to digital ocean is that, for most individuals in modern society, daily activity increasingly involves interaction with digital devices which also act as sensors in larger technology infrastructures. Multifunctional mobile computing devices (often anachronistically also called "phones") allow the unobtrusive (and sometimes unrevealed) collection and communication of data to numerous electronic aggregation points. Software embedded in the phone is often designed to capture your location in the Global Positioning System from which speed, choice of routes, and affinity for destinations can be learned. Patterns of cell phone use provide information related to social and business relationships. Accelerometers on these devices enable them to be used as game consoles and collectors of other data. An emerging practice of personal data collection is referred to as the quantified self movement (Wolf et al. 2010; Wolf 2002). In the area of medical quantified self, the early identification of a heart attack by remote examination of unobtrusive ekg data can allow for precritical treatment (Kappiarukudil and Ramesh 2010). Children at the Quest to Learn schools (Salen 2012) use digital collection techniques to track and manage their own activity and health.

While smart phones are the most common computing devices available to individuals in some countries, in many portions of the educational community, students interact primarily through general computing devices such as laptop and desktop computers. In this context, the software being used is the basis of the sensor as they are typically the data collection and management interface for the user. In such environments, the product of the interaction is often captured and stored (e.g., the document created or the outcome of the game) as well as the data of ongoing processes such as game datalogs. When working with online software through a web-browser, the bulk of non-display computing can occur on remote computers that are centrally managed for software updating as well as data collection and analysis. This intensifies the scale of data collection possible.

Within the educational world, some student segments are already shifting large portions of their educational activities into interactions with digital systems such as tutoring systems (Feng and Heffernan 2006), learning management systems (LMSs) that support online collaboration, and most recently, Massively Online Open Courses (MOOCs) (Daniel 2012). These environments are typically designed with digital instrumentation in mind in order to support learning and personalization as well as the use of learning analytics (Siemens and Long 2011) to support administrative functions.

These technological shifts in sensing, however, would be of little concern if it were not for concomitant shifts in levels of use of digital devices by the general

public and the dramatic movement towards the use of digital devices for a broad range of daily activity. These activities include social communication, entertainment, play activity, broad ranges of commerce, as well as learning for broad educational purposes including focused search and retrieve activities. One implication of these shifting patterns of activity discussed by DiCerbo and Behrens (2012) is that digital learning activities, and thereby digital learning data, are able to occur with relatively few constraints on time and location. The student who wants to learn typing or another skill during his or her "after-school" time has the opportunity to access a broad range of open educational resources (OERs) that may or may not collect or transmit data. Likewise, the use of many "informal" online activities is suggested to have positive learning outcomes (Gee 2003). While it was always well known that students read and learn outside the classroom, and that there are positive educational aspects of many "informal" activities (e.g., team sports), the ability to collect, store, and analyze records of this activity for subsequent research on these genres of activity suggest a unification of understanding activity and a breaking down of pre-digital boundaries between activity clusters. For example, while the concept of homework has always been fluid (e.g., sometimes it can be done in school), the fact that it can be done at anytime in any place using network connected computers raises the question of whether the distinction between "home" and "school" work still has much value. Likewise, a student playing an educational game (or a game with educational impact) might obtain proficiency in curricular objectives (thereby relating to the activity as a curricular object), generate and respond to assessment data (relating to it as an assessment object) and have fun and communicate to friends about performance and strategies (relating to it as a social or play object). Accordingly, DiCerbo and Behrens (2012) argue that the rise of digital devices and ubiquitous activity calls into question the conceptual boundaries that arose during the pre-digital era of the digital desert.

3.3 Conceptual Shift: Testing to Interactions

Working in the context of understanding current shifts in understanding educational assessment practices, DiCerbo and Behrens (2012) apply the language of student–system interaction from Evidence Centered Design (ECD; Mislevy et al. 2002) to understand past and current large-scale testing approaches. The delivery process described in this literature is articulated in terms of a four-process delivery model (Almond et al. 2002). While this model was originally intended to explicate assessment and tutoring system activity, subsequent analysis brought application to additional activity genres including games (Behrens et al. 2006; Shute 2011). This model suggests four core processes:

- Activity Selection: What activity is to be presented next to the learner/examinee? This process can be based on electronic student profiles or can be based on teacher's human judgment, or other methods.
- Activity Presentation/Interaction: The process of the learner/examinee interacting with an activity and obtaining data. The process could include

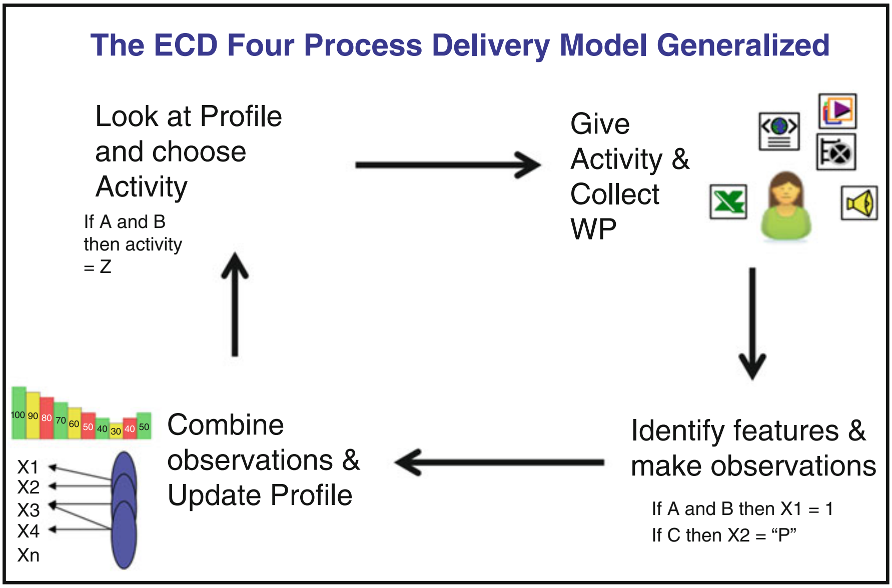

Fig. 3.1 Generalized characterization of the ECD Four-Process model following Almond et al. (2002)

answering a question or completing a complex simulation on a test, completing a level of a game, or completing a practice activity in the course of instruction. Regardless, the result is a work product that can take many forms including the answer to a question, the log of game activity, or the essay written in a project.

- Evidence Identification or Response Processing: The process of identifying observable features of the work product that can be passed to subsequent summary processes. This could include, for example, the application of Latent Semantic Analysis (Landauer et al. 1998) or other Natural Language Processing techniques to an essay that results in a list of variables with specific values. In the context of multiple choice testing, this often means the generation of a specific indicator of the correctness/incorrectness of the response. In such a context it may also be called item-level scoring.
- Evidence Accumulation or Evidence Synthesis: This is the process of summarizing previous smaller pieces of task level information to create a profile of learner states. This could be as simple as adding up all the points assigned to questions on a test to differential weighting of values based on complex statistical models such as IRT (van der Linden and Hambleton 1996) or Bayesian Inference Networks (Almond et al. 2007; Pearl 1988).

A schematic characterization of the four-process delivery model is provided in Fig. 3.1.

Table 3.1 Key differentiators between item and activity paradigm from Behrens and DiCerbo (2013)

	Item paradigm	Activity paradigm
Problem formulation	Items pose questions	Activities request action
Output	Items have answers	Activities have features
Interpretation	Items indicate correctness	Activities provide attributes
Information	Items provide focused information	Activities provide multidimensional information

DiCerbo and Behrens (2012) point out that while this is a very generalized model (see also Mislevy et al. 2012) that allows for a broad range of activity, the predominant assessment paradigm of the twentieth century was as follows:

- Activity Selection: Predetermined ordering of activities in "fixed form."
- Presentation: Questions eliciting fixed responses.
- Evidence Identification: Matching of fixed response against fixed answer.
- Evidence Synthesis: Add up "correct" responses or differentially weight them using pre-calibrated statistical models.

Let's contrast this against characterization against a similar analysis of game construction (Behrens et al. 2006):

- Activity Selection: Choose next activity or level based on state of student model.
- Presentation: May be simple or complex, possibly providing complex emulation of real or imaginary worlds.
- Evidence Identification: May be simple or complex, possibly considering strategy use, skill trajectory, and social interactions.
- Evidence Synthesis: May be simple or complex possibly using complex statistical models that may change over time.

3.3.1 Items to Activities

Behrens and DiCerbo (2013) contrasted two ends of an assessment continuum by characterizing end points of an "Item Paradigm" and an "Activity Paradigm" (see Table 3.1). The Item Paradigm is associated with relatively focused tasks that are designed to constrain the scope of possible inferences from the observations. Typically, the task is also constrained to support scalable fixed response features such as multiple choice formats. DiCerbo and Behrens (2012) argue that the technological constraints of the digital desert were a major factor in the dominance of fixed response tasks (and thereby the item paradigm) during that time. This also led to psychometric practices optimized on these formats and the corresponding constraint of the presentation processes to align with the restricted response scoring.

The activity paradigm starts with the assumption that in the new digital age, the facilities for presentation and evidence identification are not, and should not be, a

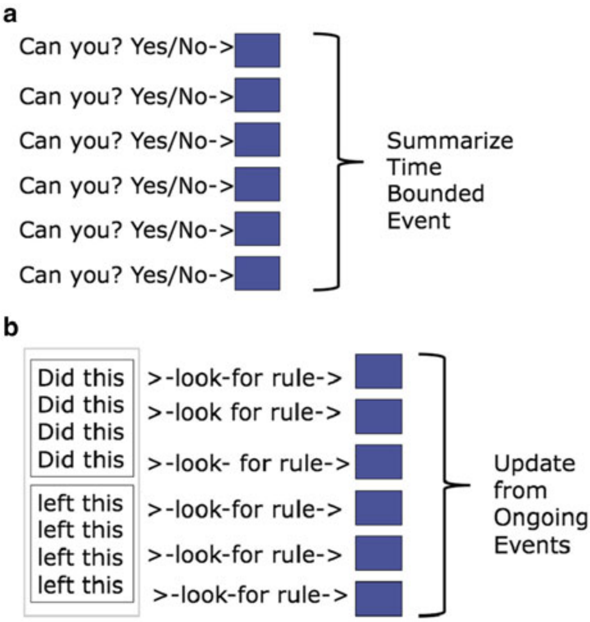

Fig. 3.2 (**a**) Characterization of the matching process in fixed response point-in-time assessment leading to summary scores, (**b**) characterization of generalized feature extraction process based on complex activity over time

primary constraint. By conceptualizing the assessment process as a feature extraction process from an activity (that may be constrained to fixed response but does not have to be), this conceptual model opens the possibility of assessment data coming from a broad range of systems including simulation-based assessment (Frezzo et al. 2009), online tutors (Feng and Heffernan 2006) or other contexts that were perhaps not originally intended to serve assessment or instructional purposes (DiCerbo 2014). These systems include activities that may ask learners to configure computer networks, engage in stepwise completion of math problems, or rescue a friend from a game-embedded vampire.

The flexibility of the four-process model is related to our ability to conceptualize and work in the activity paradigm. If we conceptualize the quantification process of measurement as linked to identifying a series of specifically constrained answers (whether on a test, survey or scoring rubric, as in Fig. 3.2a) then we have approached the problem with restricting limits to begin with and are likely to be driven increasingly toward the item paradigm. However, if we conceptualize the process as one of the feature identification from a work product (as in Fig. 3.2b), then we have new, but less bounded problems. We are free to think of user activity as a complex stream from which we seek to observe certain attributes by applying observational rules over time, rather than a set of questions that should be scored for correctness. Of course, the fixed response model in which observations are designed by the force of the fixed format, is subsumed by the more general approach of evidence identification as a process applicable to all work products. Taking this view opens up the possibility of complex scoring of activities in games (DiCerbo 2014; Shute and Ventura 2013), simulations (Frezzo, DiCerbo, Behrens

and Chen 2014), and ongoing system interaction across a number of attributes and behaviors, such as “gaming the system” (Baker et al. 2008).

We may consider an additional extension of the conceptualization discussed by DiCerbo and Behrens (2012) with regards to the “presentation” process itself. Given the original context of assessment, using the term “presentation” and extending it to activity that “requests action” is an appropriate framing. However, in attempting to expand the logic to a broader range of action, we may think not only about activities requesting action as in a test, but activities as interactions that invite, encourage, demand, attract, or otherwise motivate action and thought. To the degree that assessment becomes based on observation of natural activity, the full range of purposes and contexts of activity, and the triggers of activity should be considered. Accordingly the presentation process may be rightly renamed as an interaction or creation process given that the data-based and evidentiary outcome is a newly created work product. Shifting this language from presentation (something the delivery system does) to interaction or creation (something the learner does) opens up new possibilities for metaphor, focus, and sense making regarding the activity of the learner.

This view recommends a shift in the notion of features as fixed properties of tasks to a notion of features as emergent properties of interactions that may vary from individual to individual as different paths of action and creation provide different kinds of work products (play products? social products?) in complex systems. In the digital desert, tasks and target features need to be highly constrained for evidentiary sense making but in data-rich environments forming the digital ocean, emergent features can be detected and combined in real time.

3.4 Data Shift: Ubiquity, Connectedness, Persistence

The topics above have focused on the human activity that generates data to create the new digital ocean as well as the conceptual activity lens that we may use to understand the assessment/instruction/interaction process as it relates to data generation (presentation/interaction), transformation (evidence identification), and synthesis (evidence accumulation). In this section some of the affordances of data collection, storage, and transfer are discussed. In this regard, the chapter touches upon ubiquity, persistence, and interconnection and contrasts these new attributes of data-based systems between digital desert and digital ocean scenarios. These characteristics of data, rather than the simple generation of more data, are likely to lead to the transformational changes of the digital ocean.

3.4.1 Ubiquity

As envisioned by DiCerbo and Behrens (2012), the digital ocean exists because the shift to ever-increasing natural interaction with sensor-embedded devices allows the naturalistic and unobtrusive generation and collection of data. In the digital desert,

data collection was expensive and dedicated resources and methods needed to be employed to collect and access the requisite data. In the digital ocean, data is being generated throughout the day by human involvement with a myriad of digital systems. As those authors wrote

> This is the vision of a world in which the natural instrumentation of a digital ocean blurred the distinctions between formative and summative assessment, curriculum and assessment, and formal and informal aspects of instruction. It is a world in which data are a side effect, not the primary goal of interesting and motivating activity, and perhaps a world where 'testing' is a rare event, but assessment is 'in the water.' (DiCerbo and Behrens 2012, p. 302)

This shift is essential to understand for the field of learning analytics. The availability of data from learning-in-progress should fundamentally change how we understand and provide feedback about learning. In the digital desert, the information available to those outside the classroom came from scores on assessments that were separated from learning. Even when scores on quizzes, chapter tests, and benchmark assessments were available, these seldom provided information about the learning process, misconceptions, or other information that would assist in making decisions about the correct learning path for students. With data becoming available from a vast swath of learner activity, including activity stream information, we can now use not only the summary of the final product but information about the learners' process and progression to build models of learning.

3.4.2 Persistence

Persistence is another transformative characteristic of data in the age of the digital ocean. Persistence is important for several reasons. First, persistent data supports automated learning and decision-making over time. In the pre-digital era only gross summaries of achievement are frequently passed outside the classroom. Digital interactions make records of student experience more portable and sharable. Teachers, administrators, and parents can more easily share and inspect the artifacts of learning in this milieu. Contemporary data-driven systems may keep track of a myriad of pieces of data from multiple levels of educational environments throughout the course of student learning trajectories. This allows comparison of performance over time, the identification of patterns of behavior and knowledge acquisition over time, and the comparison of groups between each other at different points in time. Within intelligent systems, the ability to provide customized instruction depends on the existence of data persisting across time so that student profiles can be accessed in order for the customization to be appropriate. While it seems, in some ways, trivial and straightforward in the digital age, it will represent a dramatic shift in the study and facilitation of instruction as the availability of persistent data becomes commonplace in the school.

Second, persistent information may lead to improved knowledge models and research over time. The progressions of proficiency over time, learning trajectories,

and paths to overcome misconceptions described above can inform not only learners and teachers, but also learning scientists seeking to improve learning offerings.

Of course, the persistent nature of data raises many questions around privacy and data ownership, which often outstrip our current policies and regulations. These issues need to be addressed in reasonable ways that protect individuals while offering access to the potential of intelligent data availability to improve learning and instruction.

3.4.3 Interconnection

The digital ocean offers the promise of interconnection of data collected across systems. To accomplish this goal, data collected from the sundry devices will need to be linked in a way that it currently is not. Because of the evolutionary nature of technology in education, it is not uncommon for systems to be built separately for curricular or assessment data or formative and summative assessment systems. Systems designed this way fail to recognize the flexibility of the activity delivery framework and fail to take advantage of multidimensional linkages that may reveal important insights regarding patterns of learning. Group level summaries of one system and group level summaries of another system fail to reveal the interactional effects that happen across variables and individuals. In the shorter term, individual systems will be built with internal linkages that preserve the user agreements and hopefully serve end-users.

In the longer term, the systems would be able to communicate amongst themselves. Today, for example, intelligent digital tutors generally start with no information about the student. As a result, initial activity time is spent in an effort to find the correct level for the student. If other software systems shared information with the tutor, the learner could begin with information about skill levels, interests, and goals in the system and jump immediately into appropriate activity.

3.5 Corresponding Shifts for Research and Training

The shift from digital desert to digital ocean will have impacts on educational research that we believe will be dramatic, rapid, and likely difficult to anticipate at present. During the dominance of the digital desert, data per se was a seldom used concept and easily overshadowed by canonical manipulative processes such as statistics. In fact, when first attempting to name a chapter as "Data and Data Analysis," Behrens and Smith (1996) were initially told that data was not a proper subject of a methodological paper and that the title should emphasize other aspects of the analytic endeavor. In the current age of impressively increasing amounts and types of data, the opposite situation is emerging. Namely, that the enthusiasm about data, its uniqueness, volume, and corresponding display oftentimes mask discussion of the

fundamental analytic mechanics of inference that connect data to knowledge. Here we address the impact of new forms of data and analytics on scientific theory development, new tools required to facilitate such understanding, and new processes that should be considered given the new forms of data and analysis.

3.6 The Digital Ocean Changes Our Relationship to Theory

In the digital desert the relationship between the research process and the data collection process is highly constrained by the cost of data. Prior and present practice is to progress through a funnel of increasingly restricted ranges of data to be considered relevant as shown in Fig. 3.3.

While this analysis may appear cynical, it is consistent with the long-standing complaint that much institutional research is conducted on university undergraduates in laboratory conditions, not because they are the most appropriate subject pool but rather because they are the most “affordable” data source given institutional support for the process (Gordon et al. 1986). Moving forward, we think the rise in new data of the digital ocean will help answer existing questions with new frequency and depth, while at the same time producing new types of data that provoke new questions about emerging digital experiences.

3.6.1 New Data for Existing Theories

It is important to recognize how the advancement of technology has already dramatically increased the speed at which research is conducted and results are assembled and disseminated. Prior to the last few decades, the vast majority of social science research was done on paper using forms constructed by combinations of type-written or draftsman-drawn copy, later followed by bubble-sheets and, of course, most recently the computer. The amount of time required for the hand assembly, and collating of a paper survey or coding sheet would be shocking to those who never knew life before spreadsheets and personal computers.

The current early days of the digital ocean provide a stark contrast already. The prevalence of LMSs and Student Information Systems (SISs) offers the possibility of relatively easy (and unobtrusive) access to electronic data from across the university campus, whether physical, virtual, or blended. Insofar as software formalizes and automates actions, it reflects the underlying theory of action of the creators and customers. Accordingly, concerns regarding the management of education are often key goals and thereby the data collected and the analyses undertaken in current systems largely reflects mainstream concerns of educational management, often focused on computerizing pre-digital activities.

The rising digital ocean also consists of new data specifically designed to fill in gaps in data collection left by the need for large-scale and longitudinal consistency. In the educational community in the United States, the National Center for

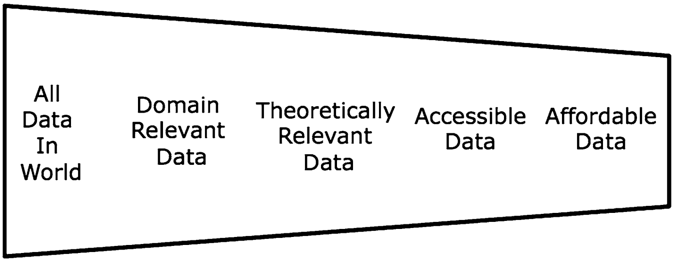

Fig. 3.3 Funnel of data constraints on scientific data

Educational Statistics (NCES; nces.gov) sponsors the data collection and dissemination of a wide variety of information at a number of levels of the educational ecosystem. Sample "products" include the National Assessment of Educational Progress (NAEP), the National Assessments of Adult Literacy (NAAL), the Educational Longitudinal Studies (ELS), National Household Education Survey (NHES), the School Crime Survey (SCS), the Library Statistics Program, and numerous other datasets. Other countries and cooperative unions such as the Organization for Economic and Cooperative Development (OECD, oecd.org/statistics) or the Interuniversity Consortium for Political and Social Research (ICPSR) provide similar access to a broad range of data and analyses. Such efforts have parallels in other disciplines such as health (e.g., Healthdata.gov) and physical sciences as well.

Considering the filter model shown in Fig. 3.3, removal or dramatic release of constraints on the right side of the figure implies a shift from data "because it's the only thing we can get" to data more aligned with the theoretical goals. For example, the data produced by the NCES discussed above has allowed the examination of a broad range of questions previously unanswerable by the single small research project funded over a short period of time.

3.6.2 New Data from New Experiences Raising New Questions

In the examples above, technology is being used largely to collect data following existing pre-digital paradigms such as surveys and tests. However, the digital shift creates both new digital experiences and new digital data. Consider social communication through methods such as the use of Twitter (twitter.com). Twitter relies largely on the availability of mobile devices and is a relatively new phenomenon raising new social issues regarding the social norms and possibilities for instantly distributed communication. Because it is a digital-from-birth technology, it is also naturally instrumented for data sharing and analytics. Twitter provides access to some of their large data corpus to the public on a daily basis through readily available computer-based application program interfaces (APIs). Computer tools for the extraction and visualization of this data are available to easily analyze and interpret some aspects of the data (Russell 2011). Similarly Google (Google.com) provides

online analytical tools to access results regarding search activity of their customers as well as basic word count analytics on their scanned text project.

New educational understandings are likewise being extended by new digital activities. We see this, for instance, in the literature in educational assessment where some researchers have shifted focus from optimization of long-established learning and assessment methods with primary focus on skill proficiency to new foci on motivation and intrinsic interest (e.g., Shute and Ventura 2013) and the integration of data-cross complex environments (Mislevy et al. 2014) such as SimCity.edu (simcityedu.org). Likewise, within the intelligent tutoring literature, the availability of large amounts of learning data is now being complimented with sample-based data addressing more difficult human attributes (Baker et al. 2008). Supplementing the large automatically collected database with sample-based data collection from the classroom, these authors were able to address complex, long-term inferences in relatively efficient ways.

3.6.3 Availability Bias

As we noted at the start of this section, researchers have always worked with the data that they have, even when it is not always the data that they want. Moving forward, we believe that the rapid changes in new types and amounts of data will make many types of data collection easier, but some more so. Researchers (and citizens) need to be on guard to consider how the availability of new forms of data bias their work in certain directions and consider what theoretical side effects arise from certain types of work.

3.7 The Digital Ocean Changes the Tools We Need

To leverage the new opportunities of digital data, analysts and researchers will require new skills and tools along with new ways to think about their endeavors. These new tools will need to respond to a number of cross-disciplinary shifts that reflect the breadth of impact that our current technology evolution is having on society and the study of its features.

3.7.1 Implications of the Shift Toward More Human Interaction with Digital Devices

Human activity, both personal and social, will increasingly be facilitated by human interaction with electronic devices. Accordingly, students should have basic literacy in the understanding of Human Computer Interaction as a frame for research as well as literacy in modern software programming tools.

Methods for studying Human Computer Interaction have evolved greatly in the last 20 years along with the rise of human–machine interaction. While the moniker of "computer" is dated as is "machine," the key idea is that there are principles of interaction analysis that can be brought to bear broadly in understanding human activity. For example, Human Computer Interaction frames used in the Computer Supported Collaborative Learning literature to illustrate Activity Theory (Engeström et al. 1999) can be applied to understand the human interactional dynamics of simulation-based assessment and instruction. While this is often an embedded view in the learning sciences literature, it is not universally built into graduate study in education.

3.7.2 *Shift from Centrality of Mathematics to Centrality of Software*

Following Seidel & Deift (2011), we consider software "the ubiquitous modern language of science." Understanding the logic of computing, the possibilities of modern applied computing, and having facility for generic data manipulation and system interaction are essential skills for modern researchers and data analysts. Recently the freely available and rapidly expanding Python language has emerged as a notable powerful and increasingly common tool for data visualization (Rossant 2013; Vaingast 2009), natural language processing (Bird et al. 2009; Perkins 2010), general data analysis (Janert 2010; McKinney 2012), and statistical manipulation (Conway and White 2012; Russell 2011). The R language (r-project.org) is likewise emerging as a widely used tool for data science though its statistical beginnings make it more appropriate for that arena than for solving universal computing problems.

Even if students are not going to obtain proficiency in a programming language, it is essential that they understand the basic logic of computing and trends in scientific computing. As a general overview and directions are needed for many researchers, this may be a ripe area for research-supporting agencies to promote professional development.

Another way to help students prepare for the emerging digital ocean is to develop familiarity with standards for data description, movement, and use as embodied in standards for computer design and data exchange. For example, the Question and Testing Interoperability specification of the IMS (QTI; IMS 2006) is an industry standard for assessment delivery. As a standard, it represents some level of consensus of practitioners in a field and represents the mental models prevalent at the time. Indeed QTI was strongly influenced by the four-process model described above, including specification of response processing and presentation processes. Other standards exist for other domains such as the predictive model markup language (PMML; Guazzelli et al. 2012) used in data mining and related statistical disciplines.

As software tools converge in feature sets (even if only in discourse), points of activity across communities of educational practice, research, and development, it is increasingly important that training and research programs understand the current practice-based conceptualizations as represented in those systems.

3.7.3 Shift from Small Sample Statistics to Large and Combined Data

The most prominent statistical frameworks of the last 100 years centered primarily around the problem of inferring population parameters from small samples (Behrens and Smith 1996). Given the current move toward data access to complete population, some common practices from these existing statistical frameworks applied to large data can be misleading. For example, using a traditional significance test approach without considerations of effect size can actually increase inferential error (Glass 1976). With large sample sizes, even very small effect sizes yield statistically significant hypothesis tests. Accordingly, researchers are likely to need to be reintroduced to large sample or population analytical methods as the inferential value of digital desert methods recedes.

3.7.4 Shift from Exploration Avoidant to Exploration Intensive

The absence of theory regarding detailed aspects of our data and experience has led, and will continue to lead, to an increased emphasis on the use of exploratory analytical techniques. Such activity is already evident in work undertaken using automated search and comparison over numerous datasets in the Pittsburgh Science of Learning Center's Data Shop open data and analysis infrastructure. New techniques such as learning factors analysis (LFA; Cen et al. 2006) attempt to recover series of optimally sloped learning curves across numerous combinations of possible variable combinations. While human guidance is often possible and likely preferred, large combinations of empirical results may be available compared with the number of available explanations.

While some may consider this an overly empiricist approach, it appears at present as the natural automation of concepts and tasks currently undertaken in common educational and psychological inference in which hypotheses are often relatively underspecified, leaving room for a mix of conformational and "unexpected" results given a particular testing setup (Behrens 1997; Gigerenzer 2009). Moreover, with very large amounts of data over tens or hundreds of thousands of learners, there is likely sufficient data for data exploration and hypothesis generation as well as confirmation on alternate data.

To address these issues, we recommend the teaching and practice of Exploratory Data Analysis (EDA, Tukey 1977). While a long-standing tradition in the statistical literature, it is largely overshadowed in practice by commonplace significance testing paradigms that can take an automated or even ritualistic view (Salsburg 1985). As an antidote, the EDA approach recommends the 4 Rs of revelation (visualization & statistical graphics), residuals (iterative model building and checking), re-expression (re-scaling and measurement), and resistance (using statistical methods

not effected by unusual observations. A recent summary by Behrens et al. (2012) is recommended for social scientists and includes a number of relatively recent visualization methods. Behrens (1997) provides a summary overview of the philosophical issues as well as clear examples of published analyses that went awry by failing to apply the EDA techniques.

3.7.5 Shift from Easy-to-Handle Data to Harder-to-Handle Bigger Data

The current Big Data movement (e.g., Franks 2012) has often been defined less by the social/methodological implications discussed in this paper, than by the sheer size of the data and the necessity of developing new computing tools to address it (but see Smolan and Erwitt (2012) for a compelling social view). For example, in large game data, a system may collect many millions of records of research data that cannot easily fit into individual machines or may extend the time required to complete and analyze to the point of making it untenable.

Students should be made aware of simple tools that can help resize and shape data. Tools such as SED and AWK and their derivatives allow for rapid extraction of key data from large files based on a simple query structure. Students will increasingly encounter Big Data that requires even more specialized approaches based on the specific technologies of Hadoop or Hadoop-based systems such as Pig and Hive. In addition, advanced students should be familiar with the basic emerging algorithms that are becoming commonplace patterns in modern computing. Computing for recommendation analysis based on collaborative filtering or other approaches seen in industry ("people like you bought X, you might want X"), for example, is an emerging common pattern (Ricci et al. 2011) that will eventually become part of the standard computing paradigm in education.

3.7.6 Shift from Lakes to Streams

As noted in the second section above, we believe a key hallmark of the emerging digital ocean is the increase in open-form data that reflects the unstructured nature of human activity. This shift requires the acquisition and application of the conceptual tools of the Four-Process Delivery model described above. These conceptual tools allow researchers to see beyond traditional data collection modes and give them a language around scientific discourse in educational domains.

In addition to the conceptual lens, students will also need to learn to compute and analyze data that is stream and event based. While this is an area of rich activity in some disciplines, advances with these types of data sources are only beginning to emerge in education.

3.8 Changes in the Research Process

3.8.1 *Shift from Research as Event to Research as an Embedded Ongoing Activity*

Reeves (2001) characterized rare summative assessments as autopsies that characterize failure after the fact and formative exams as periodic check-ups that allow for change of course before the critical condition is reached. While we agree with these characterizations, we believe that many of the digital interactions that are emerging can extend the analogy even further. From our perspective, the rise of the digital ocean as ubiquitous and unobtrusive data accumulation suggest the analogy of a heart monitor, especially when the machine is sending wireless telemetry to provide ubiquitous, unobtrusive assessment. Table 3.2 summarizes the key characteristics.

One interesting implication of this model is that granularity of information is highly correlated with the sampling temporal frequency. The table reflects a shift from data collection as a series of isolated events triggered by causes unrelated to the phenomenon being studied to an ongoing interactional model of sensor/patient monitoring and engagement. The autopsy model supposes a drop-in point-in-time researcher (coroner) is called to opportunistically take advantage of data collection opportunities. The heart monitor model assumes that there is a responsible agent in partnership with the patient to both build agency in the activity and experience of the patient as well as to support and coach the patient on the basis of increased shared information.

The ubiquity and persistence of data provide additional complexity in the methodological landscape that has been traditionally dominated in educational research by time-agnostic or time-challenged methods such as simple Analysis of Variance or repeated measures analyses limited to a relatively few data points. New datasets that may contain hundreds or thousands of data points likely require new techniques to reflect the time and dimensionality complexities.

Likewise the shift in data granularity in the digital ocean opens questions regarding whether the educational researcher plays the role of coroner or family doctor. Perhaps greater availability of data will allow the researcher to serve in a more active, continuous, supporting role while educators themselves become enabled by data access to become the newly empowered nurse practitioners. The determination of these outcomes will, in some part, depend on the evolving conceptual frames brought to the development of the devices and the human computer interactional features that evolve. It is incumbent on educational system designers to understand and study the implications of system design for learners and the stewards of learning (and research) in their ecosystems.

Table 3.2 Assessment granularity of educational artifacts and their corresponding analogs in medical examination

Educational artifact	Medical artifact
Summative end of year test	Autopsy
Formative exam	Check up
Naturalistically embedded assessment	Heart monitor

3.8.2 Shift from New Data Each Time to Ongoing Model Updating

As discussed above, given the cost of data in the digital desert, research was often conducted at single points in time. The problems of lack of replication and publication bias exacerbate this concern and are well studied in the meta-analysis literature (Glass 1976; Hedges and Olkin 1985). As standards for data collection, exchange and manipulation evolve, and access to ongoing-systems of data generation grows, there will be increased need for methodological approaches that not only describe the data at hand, but also provide an integration between new and existing data and information.

Two general approaches recommend themselves for this challenge: Bayesian statistics and meta-analysis. Meta-analysis was developed to address statistical weaknesses in the long-run and short-run probabilistic processes of significance testing, and the lack of standardization in outcomes specification leading to difficulties in comparability (Glass 1976). While meta-analysis is often conceptualized as a set of methods for summarizing a field or "conducting a literature review" the concepts and methods are amenable to many multi-results situations and recommend them even for analyzing results from within a single study if the study is heterogeneous in its goals or data. Behrens and Robinson (2005) suggested the importance of conceptualizing, analyzing, and displaying the results of multiple studies as a response surface reflecting the combined effects of the study characteristics.

Bayesian statistics have been designed precisely to take into account previously existing beliefs and conclusions and to provide a mathematical model for updating those beliefs. Accordingly, these approaches are well positioned to become a dominant paradigm in the digital ocean. In fact, the approach is currently experiencing an explosion of activity in the biological (Kery and Schaub 2011) and social sciences (Kruschke 2010) because of the computational feasibility brought about by modern computing methods (Brooks, Gelman, Jones & Meng 2011). Levy et al. (2011) provide an overview of applications of Bayesian logic to education and Gelman and Hill (2006) provide an excellent account of social science methods building on both Bayesian and Frequentist ideas.

3.9 Clarifications and Cautions

The notion of the digital ocean is not a proposal to increase instrumentation of learners for learning sake. Rather it is attempting to give voice to the possibilities embedded in the social and technological shifts that are already occurring. Digital activity is becoming commonplace in daily life and it can change how we think about assessment, learning, and education. While detailed cautions and concerns can be enumerated for many dimensions of human activity now subject to the collection of data, three broad concerns will need to suffice here.

First, the techno-social changes described in this paper and the evidence around us are poorly understood as they relate to issues of social justice and equality. Differential access to devices or intelligent computation on one's data could lead undesirable social outcomes as new types of underserved roles evolve. Educational economists and policy experts should be involved in the conversations regarding the implications of these changes for local, national, and global sociopolitical systems. This is consistent with DiCerbo and Behrens (2014) call for both "data sciences" and "data humanities."

Second, with regard to the academic/scientific communities, it is fitting to review the stories recounted in Stephen Jay Gould's *The Mismeasure of Man* (1981). While there was some controversy and disagreement over his characterization of some relatively recent scholars (e.g., disagreements regarding his characterizations of Arthur Jenson), the long historical view painted a portrait of "modern" science that requires no advanced degree to raise concern. In this volume, Gould repeatedly recounts leading scientific experts of the nineteenth and twentieth centuries coming to socially destructive and biased conclusions on the basis of seemingly new and irrefutable use of scientific data. These "objective" scientific conclusions were used to support a number of social policy conclusions that came to be considered unsubstantiated or fraudulent. Though some have argued with specific details of the book, a strong and appropriate take-away is that scientific "truths" are socially constructed and evolve over time. Translating these transient "truths" into social policy needs to be done with caution and skepticism.

Third, learning analytics, data mining, and data science are slowly emerging fields and not yet a unified discipline. Accordingly there are no clearly unified professional boundaries or long-established professional associations. While the emergence of such organizations as the Society for Learning Analytics Research (SoLAR; solaresearch.org) and the International Educational Data Mining Society (educationaldatamining.org) point toward important evolutions of a community, these groups are still emerging and lack the hallmarks of longer-established professional societies such as a common language, curricular expectations, and professional standards for ethics and scientific conduct. For example, in the educational and psychological measurement communities there are clear standards regarding the ethical standards for the construction, use, and reporting of educational and psychological tests (American Psychological Association, American Educational Research Association, and National Council on Measurement in Education 1999). Many of the principles embedded in this volume would serve as appropriate guidelines for those working in the adjacent fields of learning analytics including standards requiring professionals to consider the ability of end-users to properly understand the results of analytic activities and the responsibility of professionals to fully understand the analytic procedures used and to be able to anticipate and avoid improper use of results. Until learning analytics and data mining evolve as an integrated profession, professionals working in this area will need to seek guidance from related disciplines.

3.10 Summary/Conclusion

This paper has covered a broad range of important topics with the aim of helping the reader rethink some of their underlying assumptions regarding data and data analysis in a computationally evolving world. In the first section we established some of the privileged properties of new data and its intersection with software and hardware to enable new experiences: ubiquity, persistence, and interconnection. This suggests that we are only at the beginning of a new golden age of data, computers, and science.

Having established some sense for the potential of the computational shifts we see, we then turned to discuss a language regarding the interaction of learners and systems to collect and score data, provide it to updating models, and to potentially connect those outcomes with intelligent task selection and recommendation. Among the insights in this section, we made a careful distinction between the work product that is collected from the end-user and the observations that result from pattern recognition on them. This distinction is similar to the distinction Behrens and Smith (1996) made between "the data of the phenomenon" (work product) and "the data of the analysis" (p. 949). In the new emerging systems of ongoing data collection and model updating, additional layers of data should be considered as intermediate results and updated learner profiles evolve as new data as well.

In the subsequent section we reviewed some movements toward new forms of data collection and discussed intersections between the new forms and existing and evolving theoretical and implementation paradigms. In sum, new forms of data collection are happening at both the macro and micro levels which will have both direct and indirect effects on what can be studied and therefore, how science will progress.

In the final two sections we reviewed a number of shifts in practice that occur because of the rise of the digital ocean and recommend emphases in training or thinking that may help plan for the future. In all approaches, the wisdom and subtlety of human agency remains the best technology and it must be used in careful ways to leverage new computing approaches while avoiding the many pitfalls that may lead us to error.

References

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J.-D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement, 44*(4), 341–359.

Almond, R., Steinberg, L., & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment, 1*(5). Retrieved from https://escholarship.bc.edu/ojs/index.php/jtla/article/view/1671

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224.
Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*, 131–160.
Behrens, J. T., & DiCerbo, K. E. (2013). *Technological implications for assessment ecosystems: Opportunities for digital technology to advance assessment*. Princeton, NJ: The Gordon Commission on the Future of Assessment.
Behrens, J. T., DiCerbo, K. E., Yel, N., & Levy, R. (2012). Exploratory data analysis. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology* (Research methods in psychology 2nd ed., Vol. 2, pp. 34–70). New York: Wiley.
Behrens, J. T., Frezzo, D., Mislevy, R., Kroopnick, M., & Wise, D. (2006). Structural, functional and semiotic symmetries in simulation-based games and assessments. In E. Baker, J. Dickieson, W. Wulfeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 59–80). New York, NY: Routledge.
Behrens, J. T., & Robinson, D. H. (2005). The micro and the macro in the analysis and conceptualization of experimental data. In G. D. Phye, D. H. Robinson, & J. Levin (Eds.), *Empirical methods for evaluating educational interventions* (pp. 147–173). Burlington, MA: Elsevier.
Behrens, J. T., & Smith, M. L. (1996). Data and data analysis. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 945–989). New York: Macmillan.
Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python* (1st ed.). Sebastopol, CA: O'Reilly Media.
Brooks, S., Gelman, A., Jones, G., & Meng, X. (2011). The handbook of Markov chain Monte Carlo. Boca Raton, FL: Chapman Hall/CRC.
Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Intelligent Tutoring Systems* (pp. 164–175). Retrieved from http://link.springer.com/chapter/10.1007/11774303_17
Conway, D., & White, J. M. (2012). *Machine learning for hackers* (1st ed.). Sebastopol, CA: O'Reilly Media.
Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education, 3*. Retrieved from http://www-jime.open.ac.uk/jime/article/viewArticle/2012-18/html
DiCerbo, K. (2014). Game-based measurement of persistence. *Journal of Educational Technology and Society, 17*(1), 17–28.
DiCerbo, K., & Behrens, J. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future* (pp. 273–306). Charlotte, NC: Information Age.
DiCerbo, K. E. & Behrens, J. T. (2014). *The impact of the digital ocean on education*. [white paper] London: Pearson.
Engeström, Y., Miettinen, R., & Punamäki, R.-L. (1999). *Perspectives on activity theory*. New York, NY: Cambridge University Press.
Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. *Technology, Instruction, Cognition and Learning, 3*(1/2), 63.
Franks, B. (2012). Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. Hoboken, NJ: Wiley.
Frezzo, D. C., Behrens, J. T., Mislevy, R. J., West, P., & DiCerbo, K. E. (2009). Psychometric and evidentiary approaches to simulation assessment in Packet Tracer software. In *Fifth International Conference on Networking and Services, 2009. ICNS'09* (pp. 555–560). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4976818
Frezzo, D. C., DiCerbo, K. E., Behrens, J. T., & Chen, M. (2014). An extensible micro-world for learning in the networking professions. *Information Sciences, 264*, 91–103.
Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 20.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). New York, NY: Cambridge University Press.
Gigerenzer, G. (2009). Surrogates for theory. APS Observer, 22(2). Retrieved from http://www.psychologicalscience.org/index.php/publications/observer/2009/february-09/surrogates-for-theory.html.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.
Gordon, M. E., Slade, L. A., & Schmitt, N. (1986). The "Science of the Sophomore" revisited: From conjecture to empiricism. *Academy of Management Review, 11*, 191–207.
Gould, S. J. (1981). *The mismeasure of man*. New York: W. W. Norton.
Guazzelli, A., Lin, W.-C., & Jena, T. (2012). *PMML in action: Unleashing the power of open standards for data mining and predictive analytics* (2nd ed.). Los Angeles: CreateSpace Independent Publishing Platform.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis* (1st ed.). New York, NY: Academic Press.
IMS. (2006). *IMS question & test interoperability specification v2.0/v2.1*. Retrieved March 2013, from http://www.imsglobal.org/question/index.html
Janert, P. K. (2010). *Data analysis with open source tools* (1st ed.). Sebastopol, CA: O'Reilly Media.
Kappiarukudil, K. J., & Ramesh, M. V. (2010). Real-time monitoring and detection of "heart attack" using wireless sensor networks. In *2010 Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM)* (pp. 632–636). doi:10.1109/SENSORCOMM.2010.99.
Kery, M., & Schaub, M. (2011). *Bayesian population analysis using WinBUGS: A hierarchical perspective* (1st ed.). New York, NY: Academic Press.
Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial with R and BUGS* (1st ed.). New York, NY: Academic Press.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284.
Levy, R., Mislevy, R. J., & Behrens, J. T. (2011). MCMC in educational research. In S. Brooks, A. Gelman, G. Jones, & X. L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 531–546). Boca Raton: Chapman and Hall/CRC.
McKinney, W. (2012). *Python for data analysis*. Sebastopol, CA: O'Reilly Media.
Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence centered design, psychometrics, and data mining. *Journal of Educational Data Mining, 4*(1), 11–48.
Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K. E., John, M., Bauer, M. I., et al. (2014). *Psychometrics and game-based assessment*. New York, NY: Institute of Play.
Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19*(4), 477–496.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann. Retrieved from http://books.google.com/books?hl=en&lr=&id=AvNID7LyMusC&oi=fnd&pg=PA1&dq=Pearl+bayesian+networks&ots=FY-OSfkwZ6&sig=3J1ZLPeMMUfZNG_k73CkHClSj7o
Perkins, J. (2010). *Python text processing with NLTK 2.0 cookbook*. Birmingham, UK: Packt Publishing.
Reeves, D. (2001). *101 questions and answers about standards, assessment, and accountability*. Englewood, CO: Lead+Learn Press.
Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender systems handbook*. Berlin: Springer.
Rossant, C. (2013). *Learning IPython for interactive computing and data visualization*. Birmingham, UK: Packt Publishing.
Russell, M. A. (2011). *Mining the social web: Analyzing data from Facebook, Twitter, LinkedIn, and other social media sites* (1st ed.). Sebastopol, CA: O'Reilly Media.

Salen, K. (2012). *Seminar*. Presented at the Educational Testing Service, Princeton, NJ.

Salsburg, D. S. (1985). The religion of statistics as practiced in medical journals. *The American Statistician, 39*, 220–223.

Seidel, E., & Deift, A. (2011). Data: A centuries-old revolution in science, Part II. Society for Industrial and Applied Mathematics News, 44(7). Retrieved from https://www.siam.org/news/news.php?id=1908.

Shute, V. J. (2011). *Stealth assessment in computer-based games to support learning. Computer games and instruction*. Charlotte, NC: Information Age Publishers. Retrieved from http://myweb.fsu.edu/vshute/pdf/shute%20pres_h.pdf

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge, MA: The MIT Press.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–32.

Smolan, R., & Erwitt, J. (2012). *The human face of big data* (1st ed.). Sausalito, CA: Against All Odds Productions.

Tukey, J. T. (1977). *Exploratory data analysis*. Reading, MA: Addison Wesley.

Vaingast, S. (2009). *Beginning python visualization: Crafting visual transformation scripts* (1st ed.). Berkeley, CA: Apress.

Van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*. Springer. Retrieved from http://books.google.com/books?hl=en&lr=&id=aytUuwl4ku0C&oi=fnd&pg=PR5&dq=hambleton+item+response+theory&ots=JXdX5GjwfM&sig=KKbXwfRqSFqzMLsU_2BTY-rhcFk

Wolf, G. (2002, October 5). The data driven life. *The New York Times Magazine*, pp. 38–45.

Wolf, G., Carmichael, A., & Kelly, K. (2010). The quantified self. *TED*. Retrieved from http://www.ted.com/talks/gary_wolf_the_quantified_self.html

Chapter 4
Educational Data Mining and Learning Analytics

Ryan Shaun Baker and Paul Salvador Inventado

4.1 Introduction

In this article, we will discuss a research area/community with close ties to the learning analytics community discussed throughout this book, educational data mining (EDM). This chapter will introduce the EDM community, its methods, ongoing trends in the area, and give some brief thoughts on its relationship to the learning analytics community.

EDM can be seen in two ways; either as a research community or as an area of scientific inquiry. As a research community, EDM can be seen as a sister community to learning analytics. EDM first emerged in a workshop series starting in 2005, which became an annual conference in 2008 and spawned a journal in 2009 and a society, the International Educational Data Mining Society, in 2011. A timeline of key events in the formation of the EDM community can be seen in Fig. 4.1.

As of this writing, the EDM Society has 240 paid members, and the conference has an annual attendance around the same number. Many of the same people attend both EDM and the Learning Analytics and Knowledge (LAK) conference, and the general attitude between the two conferences is one of friendly collaboration and/or friendly competition.

As an area of scientific inquiry, EDM is concerned with the analysis of large-scale educational data, with a focus on automated methods. There is considerable thematic overlap between EDM and learning analytics. In particular, both

R.S. Baker (✉)
Teachers College, Columbia University, New York, NY, USA
e-mail: baker2@exchange.tc.columbia.edu

P.S. Inventado
Center for Empathic Human-Computer Interactions, College of Computer Studies,
De La Salle University, Manila, Philippines

Architecture for Intelligence, The Institute of Scientific and Industrial Research,
Osaka University, Osaka, Japan

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_4, © Springer Science+Business Media New York 2014

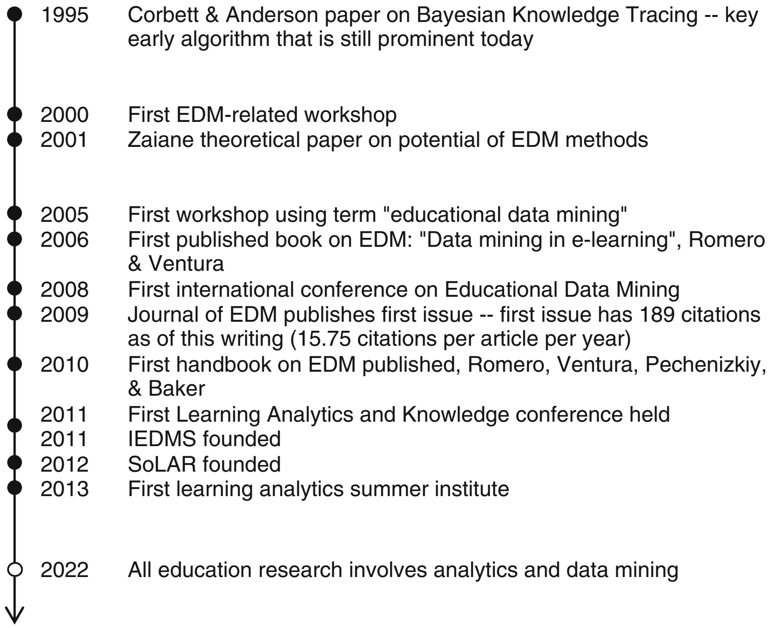

Fig. 4.1 Timeline of significant milestones in EDM

communities share a common interest in data-intensive approaches to education research, and share the goal of enhancing educational practice. At the same time, there are several interesting differences, with one viewpoint on the differences given in (Siemens and Baker 2012). In that work, it was argued that there are five key areas of difference between the communities, including a preference for automated paradigms of data analysis (EDM) versus making human judgment central (LA), a reductionist focus (EDM) versus a holistic focus (LA), and a comparatively greater focus on automated adaptation (EDM) versus supporting human intervention (LA). Siemens and Baker noted that these differences reflected general trends in the two communities rather than hard-and-fast rules. They also noted differences in preferred methodology between the two communities, a topic which we will return to throughout this chapter. Another perspective on the difference between the communities was offered in a recent talk by John Behrens at the LAK 2012 conference, where Dr. Behrens stated that (somewhat contrary to the names of the two communities), EDM has a greater focus on learning as a research topic, while learning analytics has a greater focus on aspects of education beyond learning. In our view, the overlap and differences between the communities is largely organic, developing from the interests and values of specific researchers rather than reflecting a deeper philosophical split or antagonism.

In the remainder of this chapter, we will review the key methods of EDM and ongoing trends, returning to the issue of how EDM compares methodologically to learning analytics as we do so.

4.2 Key EDM Methods

A wide range of EDM methods have emerged through the last several years. Some are roughly similar to those seen in the use of data mining in other domains, whereas others are unique to EDM. In this section we will discuss four major classes of methods that are in particularly frequent use by the EDM community, including: (a) Prediction Models, (b) Structure Discovery, (c) Relationship Mining, and (d) Discovery with Models. This is not an exhaustive selection of EDM methods; more comprehensive reviews can be found in (Baker and Yacef 2009; Romero and Ventura 2007, 2010; Scheuer and McLaren 2011). Instead, we focus on a subset of methods that are in particularly wide use within the EDM community.

4.2.1 Prediction Methods

In prediction, the goal is to develop a model which can infer a single aspect of the data (the predicted variable, similar to dependent variables in traditional statistical analysis) from some combination of other aspects of the data (predictor variables, similar to independent variables in traditional statistical analysis).

In EDM, classifiers and regressors are the most common types of prediction models, and each has several subtypes, which we will discuss below. Classifiers and regressors have a rich history in data mining and artificial intelligence, which is leveraged by EDM research. The area of latent knowledge estimation is of particular importance within EDM, and work in this area largely emerges from the User Modeling, Artificial Intelligence in Education, and Psychometrics/Educational Measurement traditions.

Prediction requires having labels for the output variable for a limited dataset, where a label represents some trusted ground truth information about the predicted variable's value in specific cases. Ground truth can come from a variety of sources, including "natural" sources such as whether a student chooses to drop out of college (Dekker et al. 2009), state-standardized exam scores (Feng et al. 2009), or grades assigned by instructors, and in approaches where labels are created solely to use as ground truth, using methods such as self-report (cf. D'Mello et al. 2008), video coding (cf. D'Mello et al. 2008), field observations (Baker et al. 2004), and text replays (Sao Pedro et al. 2010).

Prediction models are used for several applications. They are most commonly used to predict what a value will be in contexts where it is not desirable to directly obtain a label for that construct. This is particularly useful if it can be conducted in real time, for instance to predict a student's knowledge (cf. Corbett and Anderson 1995) or affect (D'Mello et al. 2008; Baker et al. 2012) to support intervention, or to predict a student's future outcomes (Dekker et al. 2009; San Pedro et al. 2013). Prediction models can also be used to study which specific constructs play an important role in predicting another construct (for instance, which behaviors are associated with the eventual choice to attend high school) (cf. San Pedro et al. 2013).

4.2.1.1 Classification

In classifiers, the predicted variable can be either a binary or categorical variable. Some popular classification methods in educational domains include decision trees, random forests, decision rules, step regression, and logistic regression. In EDM, classifiers are typically validated using cross-validation, where part of the dataset is repeatedly and systematically held out and used to test the goodness of the model. Cross-validation should be conducted at multiple levels, in line with what type of generalizability is desired; for instance, it is typically standard in EDM for researchers to cross-validate at the student level in order to ensure that the model will work for new students, although researchers also cross-validate in terms of populations or learning content. Note that step regression and logistic regression, despite their names, are classifiers rather than regressors. Some common metrics used for classifiers include A'/AUC (Hanley and McNeil 1982), kappa (Cohen 1960), precision (Davis and Goadrich 2006), and recall (Davis and Goadrich 2006); accuracy, often popular in other fields, is not sensitive to base rates and should only be used if base rates are also reported.

4.2.1.2 Regression

In regression, the predicted variable is a continuous variable. The most popular regressor within EDM is linear regression, with regression trees also fairly popular. Note that a model produced through this method is mathematically the same as linear regression as used in statistical significance testing, but that the method for selecting and validating the model in EDM's use of linear regression is quite different than in statistical significance testing. Regressors such as neural networks and support vector machines, which are prominent in other data mining domains, are somewhat less common in EDM. This is thought to be because the high degrees of noise and multiple explanatory factors in educational domains often lead to more conservative algorithms being more successful. Regressors can be validated using the same overall techniques as that in classifiers, often using the metrics of linear correlation or root mean squared error (RMSE).

4.2.1.3 Latent Knowledge Estimation

One special case of classification that is particularly important in EDM is latent knowledge estimation. In latent knowledge estimation, a student's knowledge of specific skills and concepts is assessed by their patterns of correctness on those skills (and occasionally other information as well). The word "latent" refers to the idea that knowledge is not directly measurable, it must be inferred from a student's performance. Inferring a student's knowledge can be useful for many goals—it can be a meaningful input to other analyses (we discuss this use below, in the section on discovery with models), it can be useful for deciding when to advance a student in a curriculum (Corbett and Anderson 1995) or intervene in

other ways (cf. Roll et al. 2007), and it can be very useful information for instructors (Feng and Heffernan 2007).

The models used for estimating latent knowledge in online learning typically differ from the psychometric models used in paper tests or in computer-adaptive testing, as the latent knowledge in online learning is itself dynamic. The models used for latent knowledge estimation in EDM come from two sources: new takes on classical psychometric approaches, and research on user modeling/artificial intelligence in education literature. A wide range of algorithms exists for latent knowledge estimation. The classic algorithm is either Bayes Nets (Martin and VanLehn 1995; Shute 1995) for complex knowledge structures, or Bayesian Knowledge Tracing (Corbett and Anderson 1995) for cases where each problem or problem step is primarily associated with a single skill at the point in time when it is encountered. Recently, there has also been work suggesting that an approach based on logistic regression, Performance Factors Assessment (Pavlik et al. 2009), can be effective for cases where multiple skills are relevant to a problem or problem step at the same time. Work by Pardos and colleagues (2012) has also found evidence that combining multiple approaches through ensemble selection can be more effective for large datasets than single models.

4.2.2 Relationship Mining

In relationship mining, the goal is to discover relationships between variables in a dataset with a large number of variables. This may take the form of attempting to find out which variables are most strongly associated with a single variable of particular interest, or may take the form of attempting to discover which relationships between any two variables are strongest. Broadly, there are four types of relationship mining in common use in EDM: association rule mining, sequential pattern mining, correlation mining, and causal data mining. Association rule mining comes from the field of data mining, in particular from "market basket" analysis used in mining of business data (Brin et al. 1997); sequential pattern mining also comes from data mining, with some variants emerging from the bioinformatics community; correlation mining has been a practice in statistics for some time (and the methods of post hoc analysis came about in part to make this type of method more valid); causal data mining also comes from the intersection of statistics and data mining (Spirtes et al. 2000).

4.2.2.1 Association Rule Mining

In association rule mining, the goal is to find if-then rules of the form that if some set of variable values is found, another variable will generally have a specific value. For example, a rule might be found of the form:

IF student is frustrated **OR** has a stronger goal of learning than performance
THEN the student frequently asks for help

Rules uncovered by association rule mining reveal common co-occurrences in data which would have been difficult to discover manually. Association rule mining has been used for a variety of applications in EDM. For example, Ben-Naim and colleagues (2009) found association rules within student data from an engineering class, representing patterns of successful student performance, and Merceron and Yacef (2005) studied which student errors tend to go together.

There is ongoing debate as to which metrics lead to finding the most interesting and usable association rules; a discussion of this issue can be found in Merceron and Yacef (2008), who recommend in particular cosine and lift.

4.2.2.2 Sequential Pattern Mining

In sequential pattern mining, the goal is to find temporal associations between events. Two paradigms are seen that find sequential patterns—classical sequential pattern mining (Srikant and Agrawal 1996), which is a special case of association rule mining, and motif analysis (Lin et al. 2002), a method often used in bioinformatics to find common general patterns that can vary somewhat. These methods, like association rule mining, have been used for a variety of applications, including to study what paths in student collaboration behaviors lead to a more successful eventual group project (Perera et al. 2009), the patterns in help-seeking behavior over time (Shanabrook et al. 2010), and studying which patterns in the use of concept maps are associated with better overall learning (Kinnebrew and Biswas 2012). Sequential pattern mining algorithms, like association rule mining algorithms, depend on a number of parameters to select which rules are worth outputting.

4.2.2.3 Correlation Mining

In correlation mining, the goal is to find positive or negative linear correlations between variables. This goal is not a new one; it is a well-known goal within statistics, where a literature has emerged on how to use post hoc analysis and/or dimensionality reduction techniques in order to avoid finding spurious relationships. The False Discovery Rate paradigm (cf. Benjamini and Hochberg 1995; Storey 2003) has become increasingly popular among data mining researchers across a number of domains. Correlation mining has been used to study the relationship between student attitudes and help-seeking behaviors (Arroyo and Woolf 2005; Baker et al. 2008), and to study the relationship between the design of intelligent tutoring systems and whether students game the system (Baker et al. 2009).

4.2.2.4 Causal Data Mining

In causal data mining, the goal is to find whether one event (or observed construct) was the cause of another event (or observed construct) (Spirtes et al. 2000). Causal data mining is distinguished from prediction in its attempts to find not just

predictors but actual causal relationships, through looking at the patterns of covariance between those variables and other variables in the dataset. Causal data mining in packages such as TETRAD (Scheines et al. 1998) has been used in EDM to predict which factors will lead a student to do poorly in a class (Fancsali 2012), to analyze how different conditions of a study impact help use and learning differently (Rau and Scheines 2012), and to study how gender and attitudes impact behaviors in an intelligent tutor and consequent learning (Rai and Beck 2011).

4.2.3 Structure Discovery

Structure discovery algorithms attempt to find structure in the data without any ground truth or a priori idea of what should be found. In this way, this type of data mining contrasts strongly with prediction models, above, where ground truth labels must be applied to a subset of the data before model development can occur. Common structure discovery algorithms in educational data include clustering, factor analysis, and domain structure discovery algorithms. Clustering and factor analysis have been used since the early days of the field of statistics, and were refined and explored further by the data mining and machine learning communities. Domain structure discovery emerged from the field of psychometrics/educational measurement.[1]

As methods that discover structure without ground truth, less attention is generally given to validation than in prediction, though goodness and fit calculations are still used in determining if a specific structure is superior to another structure.

4.2.3.1 Clustering

In clustering, the goal is to find data points that naturally group together, splitting the full dataset into a set of clusters (Kaufman and Rousseeuw 1990). Clustering is particularly useful in cases where the most common categories within the dataset are not known in advance. If a set of clusters is optimal, each data point in a cluster will in general be more similar to the other data points in that cluster than the data points in other clusters. Clusters can be created at several different grain sizes. For example, schools could be clustered together (to investigate similarities and differences among schools), students could be clustered together (to investigate similarities and differences among students), or student actions could be clustered together (to investigate patterns of behavior) (cf. Amershi and Conati 2009; Beal et al. 2006). Clustering algorithms typically split into two categories: hierarchical approaches such as hierarchical agglomerative clustering (HAC), and non-hierarchical approaches such as *k*-means, gaussian mixture modeling (sometimes referred to as

[1] A fourth type of structure discovery, Network Analysis, is more characteristic of work in learning analytics than in educational data mining (cf. Dawson 2008; Suthers and Rosen 2011), and is not discussed in detail here for that reason.

EM-based clustering), and spectral clustering. The key difference is that hierarchical approaches assume that clusters themselves cluster together, whereas non-hierarchical approaches assume that clusters are separate from each other.

4.2.3.2 Factor Analysis

In factor analysis, the goal is to find variables that naturally group together, splitting the set of variables (as opposed to the data points) into a set of latent (not directly observable) factors (Kline 1993). Factor analysis is frequently used in psychometrics for validating or determining scales. In EDM, factor analysis is used for dimensionality reduction (e.g., reducing the number of variables), including in preprocessing to reduce the potential for overfitting and to determine meta-features. One example of its use in EDM is work to determine which features of intelligent tutoring systems group together (cf. Baker et al. 2009); another example is as a step in the process of developing a prediction model (cf. Minaei-Bidgoli et al. 2003). Factor analysis includes algorithms such as principal component analysis and exponential-family principal components analysis.

4.2.3.3 Domain Structure Discovery

Domain structure discovery consists of finding which items map to specific skills across students. The Q-Matrix approach for doing so is well-known in psychometrics (cf. Tatsuoka 1995). Considerable work has recently been applied to this problem in EDM, for both test data (cf. Barnes et al. 2005; Desmarais 2011), and for data tracking learning during use of an intelligent tutoring system (Cen et al. 2006). Domain structures can be compared using information criteria metrics (Koedinger et al. 2012), which assess fit compared to the complexity of the model (more complex models should be expected to spuriously fit data better). A range of algorithms can be used for domain structure discovery, from purely automated algorithms (cf. Barnes et al. 2005; Desmarais 2011; Thai-Nghe et al. 2011), to approaches that utilize human judgment within the model discovery process such as learning factors analysis (LFA; Cen et al. 2006).

4.2.4 Discovery with Models

In discovery with models, a model of a phenomenon is developed via prediction, clustering, or in some cases knowledge engineering (within knowledge engineering, the model is developed using human reasoning rather than automated methods). This model is then used as a component in a second analysis or model, for example in prediction or relationship mining. Discovery with models is not common in data mining in general, but is seen in some form in many computational science domains.

In the case of EDM, one common use is when an initial model's predictions (which represent predicted variables in the original model) become predictor variables in a new prediction model. For instance, prediction models of robust student learning have generally depended on models of student meta-cognitive behaviors (cf. Baker et al. 2011a, b), which have in turn depended on assessments of latent student knowledge (cf. Aleven et al. 2006). These assessments of student knowledge have in turn depended on models of domain structure.

When using relationship mining, the relationships between the initial model's predictions and additional variables are studied. This enables a researcher to study the relationship between a complex latent construct and a wide variety of observable constructs, for example investigating the relationship between gaming the system (as detected by an automated detector) and student individual differences (Baker et al. 2008).

Often, discovery with models leverages the generalization of a prediction model across contexts. For instance, Baker and Gowda (2010) used predictions of gaming the system, off-task behavior, and carelessness across a full year of educational software data to study the differences in these behaviors between an urban, rural, and suburban school in the same region.

4.3 Trends in EDM Methodologies and Research

Given that "educational data mining" has been around as a term for almost a decade at this writing, and several early EDM researchers had been working in this area even before the community had begun to coalesce, we can begin to see trends and changes in emphasis occurring over time.

One big shift in EDM is the relative emphasis given to relationship mining. In the early years of EDM, relationship mining was used in almost half of the articles published (Baker and Yacef 2009). Relationship mining methods have continued to be important in EDM since then, but it is fair to say that the dominance of relationship mining has reduced somewhat in the following years. For example in the EDM2012 conference, only 16 % of papers use relationship mining as defined in this article.

Prediction and clustering were important methods in the early years of EDM (Baker and Yacef 2009), and have continued to be highly used. However, within the category of prediction modeling, the distribution of methods has changed substantially. Classification and regression were important in 2005–2009, and remain important to this day, but latent knowledge estimation has increased substantially in importance, with articles representing different paradigms for how to estimate student knowledge competing to see which algorithms are most effective in which contexts (Pavlik et al. 2009; Gong et al. 2011; Pardos et al. 2012).

A related trend is the increase in the prominence of domain structure discovery in recent EDM research. Although domain structure discovery has been part of EDM from the beginning (Barnes 2005), recent years have seen increasing work on

a range of approaches for modeling domains. Some work has attempted to find better ways to find q-matrices expressing domain structure in a purely empirical fashion (Desmarais 2011; Desmarais et al. 2012), while other work attempts to leverage human judgment in fitting q-matrices (Cen et al. 2007; Koedinger et al. 2012). Additionally, in recent years there has been work attempting to automatically infer prerequisite structures in data (Beheshti and Desmarais 2012), and to study the impact of not following prerequisite structures (Vuong et al. 2011).

A third emerging emphasis in EDM is the continued trend towards modeling a greater range of constructs. Though the trends in latent knowledge estimation and domain structure discovery reflect the continued emphasis within EDM on modeling student knowledge and skill, there has been a simultaneous trend towards expanding the space of constructs modeled through EDM, with researchers expanding from modeling knowledge and learning to modeling constructs such as metacognition, self-regulation, motivation, and affect (cf. Goldin et al. 2012; Bouchet et al. 2012; Baker et al. 2012). The increase in the range of constructs being modeled in EDM has been accompanied by an increase in the number of discovery with models analyses leveraging those models to support basic discovery.

4.4 EDM and Learning Analytics

Many of the same methodologies are seen in both EDM and Learning Analytics. Learning analytics has a relatively greater focus on human interpretation of data and visualization (though there is a tradition of this in EDM as well—cf. Kay et al. 2006; Martinez et al. 2011). EDM has a relatively greater focus on automated methods. But ultimately, in our view, the differences between the two communities are more based on focus, research questions, and the eventual use of models (cf. Siemens and Baker 2012), than on the methods used.

Prediction models are prominent in both communities, for instance, although Learning Analytics researchers tend to focus on classical approaches of classification and regression more than on latent knowledge estimation. Structure Discovery is prominent in both communities, and in particular clustering has an important role in both communities. In terms of specialized/domain-specific structure discovery algorithms, domain structure discovery is more emphasized by EDM researchers while network analysis/social network analysis is more emphasized in learning analytics (Bakharia and Dawson 2011; Schreurs et al. 2013), again more due to research questions adopted by specific researchers, than a deep difference between the fields. Relationship mining methods are significantly more common in EDM than in learning analytics. It is not immediately clear to the authors of this paper why relationship mining methods have been less utilized in learning analytics than in EDM, given the usefulness of these methods for supporting interpretation by analysts (this point is made in d'Aquin and Jay, 2013, who demonstrate the use of sequential pattern mining in learning analytics). Discovery with models is significantly more common in EDM than learning analytics, and much of its appearance at LAK

conferences is in papers written by authors more known as members of the EDM community (e.g., Pardos et al. 2013). This is likely to again be due to differences in research questions and focus; even though both communities use prediction modeling, LAK papers tend to predict larger constructs (such as dropping out and course failure) whereas EDM papers tend to predict smaller constructs (such as boredom and short-term learning), which are more amenable to then use in discovery with analyses of larger constructs.

Finally, some methodological areas are more common in learning analytics than in EDM (though relatively fewer, owing to the longer history of EDM). The most prominent example is the automated analysis of textual data. Text analysis, text mining, and discourse analysis is a leading area in learning analytics; it is only seen occasionally in EDM (cf. D'Mello et al. 2010; Rus et al. 2012).

4.5 Conclusion

In recent years, two communities have grown around the idea of using large-scale educational data to transform practice in education and education research. As this area emerges from relatively small and unknown conferences to a theme that is known throughout education research, and which impacts schools worldwide, there is an opportunity to leverage the methods listed above to accomplish a variety of goals. Every year, the potential applications of these methods become better known, as researchers and practitioners utilize these methods to study new constructs and answer new research questions.

While we learn where these methods can be applied, we are also learning how to apply them more effectively. Having multiple communities and venues to discuss these issues is beneficial; having communities that select work with different values and perspectives will support the development of a field that most effectively uses large-scale educational data. Ultimately, the question is not which methods are best, but which methods are useful for which applications, in order to improve the support for any person who is learning, whenever they are learning.

References

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*(2), 101–128.

Amershi, S., & Conati, C. (2009). Combining unsupervised and supervised classification to build user models for exploratory learning environments. *Journal of Educational Data Mining, 1*(1), 18–71.

Arroyo, I., & Woolf, B. (2005). Inferring learning and attitudes from a Bayesian Network of log file data. In: *Proceedings of the 12th International Conference on Artificial Intelligence in Education* (pp. 33–40).

Baker, R., Corbett. A. T., Koedinger, K., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students game the system. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383–390).

Baker, R., de Carvalho, A., Raspat, J., Aleven, V., Corbett, A., & Koedinger, K. (2009). Educational software features that encourage and discourage “gaming the system”. In: *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 475–482).

Baker, R., & Gowda, S. (2010). An analysis of the differences in the frequency of students’ disengagement in urban, rural, and suburban high schools. In: *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 11–20).

Baker, R., Gowda, S. M., & Corbett, A. T. (2011a). Towards predicting future transfer of learning. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial intelligence in education*: *Vol. 6738. Lecture notes in computer science* (pp. 23–30). Heidelberg, Germany: Springer.

Baker, R., Gowda, S. M., & Corbett, A. T. (2011b). Automatically detecting a student’s preparation for future learning: Help use is key. In *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 179–188).

Baker, R., Kalka, J., Aleven, V., Rossi, L., Gowda, S., Wagner, A., et al. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 126–133).

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in “gaming the system” behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224.

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Bakharia, A., & Dawson, S. (2011). SNAPP: A bird’s-eye view of temporal participant interaction. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 168–173).

Barnes, T. (2005). The q-matrix method: Mining student response data for knowledge. In: *Proceedings of the American Association for Artificial Intelligence 2005 Educational Data Mining Workshop* (pp. 39–46).

Barnes, T., Bitzer, D., & Vouk, M. (2005). Experimental analysis of the q-matrix method in knowledge discovery. In M.-S. Hacid, N. Murray, Z. Raś, & S. Tsumoto (Eds.), *Foundations of intelligent systems: Vol. 3488. Lecture notes in computer science* (pp. 603–611). Heidelberg, Germany: Springer.

Beal, C. R., Qu, L., & Lee, H. (2006). Classifying learner engagement through integration of multiple data sources. In: *Proceedings of the 21st National Conference on Artificial Intelligence* (pp. 151–156).

Beheshti, B., & Desmarais, M. (2012). Improving matrix factorization techniques of student test data with partial order constraints. In: *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization* (pp. 346–350).

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.

Ben-Naim, D., Bain, M., & Marcus, N. (2009). A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials. In: *Proceedings of the 2nd International Conference on Educational Data Mining* (pp. 21–30).

Bouchet, F., Azevedo, R., Kinnebrew, J., & Biswas, G. (2012). Identifying students’ characteristic learning behaviors in an intelligent tutoring system fostering self-regulated learning. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 65–72).

Brin, S., Motwani, R., Ullman, J., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. In: *Proceedings of the 1997 ACM International Conference on Management of Data* (pp. 255–264).

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis—A general method for cognitive model evaluation and improvement. In M. Ikeda, K. Ashley, & T.-W. Chan (Eds.), *Intelligent tutoring systems: Vol. 4053. Lecture notes in computer science* (pp. 164–175). Heidelberg, Germany: Springer.

Cen, H., Koedinger, K., & Junker, B. (2007). Is over practice necessary?—Improving learning efficiency with the cognitive tutor through educational data mining. In: *Proceedings of 13th International Conference on Artificial Intelligence in Education* (pp. 511–518).

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Corbett, A., & Anderson, J. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253–278.

d'Aquin, M., & Jay, N. (2013). Interpreting data mining results with linked data for learning analytics: Motivation, case study and directions. In: *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 155–164).

D'Mello, S., Craig, S., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1–2), 45–80.

D'Mello, S., Olney, A., & Person, N. (2010). Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining, 2*(1), 1–37.

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233–240).

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology and Society, 11*(3), 224–238.

Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting students drop out: A case study. In: *Proceedings of 2nd International Conference on Educational Data Mining* (pp. 41–50).

Desmarais, M. (2011). Conditions for effectively deriving a q-matrix from data with non-negative matrix factorization. In: *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 41–50).

Desmarais, M., Beheshti, B., & Naceur, R. (2012). Item to skills mapping: Deriving a conjunctive q-matrix from data. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K.-K. Panourgia (Eds.), *Intelligent tutoring systems: Vol. 7315. Lecture notes in computer science* (pp. 454–463). Heidelberg, Germany: Springer.

Fancsali, S. (2012). Variable construction and causal discovery for cognitive tutor log data: Initial results. In: *Proceedings of the 5th Conference on Educational Data Mining* (pp. 238–239).

Feng, M., & Heffernan, N. (2007). Towards live informing and automatic analyzing of student learning: Reporting in the assistment system. *Journal of Interactive Learning Research, 18*(2), 207–230.

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction, 19*(3), 243–266.

Goldin, I., Koedinger, K. R., & Aleven, V. (2012). Learner differences in hint processing. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 73–80).

Gong, Y., Beck, J. E., & Heffernan, N. T. (2011). How to construct more accurate student models: Comparing and optimizing knowledge tracing and performance factor analysis. *International Journal of Artificial Intelligence in Education, 21*(1), 27–46.

Hanley, A., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Kay, J., Maisonneuve, N., Yacef, K., & Zaïane, O. (2006). Mining patterns of events in students' teamwork data. In: *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems* (pp. 45–52).

Kinnebrew, J., & Biswas, G. (2012). Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 57–64).

Kline, P. (1993). *An easy guide to factor analysis*. London: Routledge.

Koedinger, K., McLaughlin, E., & Stamper, J. (2012). Automated student model improvement. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 17–24).

Lin, J., Keogh, E., Lonardi, S., & Patel, P. (2002). Finding motifs in time series. In: *Proceedings of the 2nd Workshop on Temporal Data Mining* (pp. 53–68).

Martin, J., & VanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human Computer Studies, 42*(6), 575–592.

Martinez, R., Yacef, K., Kay, J., Kharrufa, A., & Al-Qaraghuli, A. (2011). Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In: *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 111–120).

Merceron, A., & Yacef, K. (2005). Educational data mining: A case study. In: *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning Through Socially Informed Technology* (pp. 467–474).

Merceron, A., & Yacef, K. (2008). Interestingness measures for association rules in educational data. In: *Proceedings of the 1st International Conference on Educational Data Mining* (pp. 57–66).

Minaei-Bidgoli, B., Kashy, D., Kortmeyer, G., & Punch, W. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. In: *Frontiers in Education, 2003. FIE 2003 33rd Annual* (pp. T2A 13–18). (http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1263284&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F8925%2F28250%2F01263284.pdf%3Farnumber%3D1263284#).

Pardos, Z., Baker, R., San Pedro, M., Gowda, S., & Gowda, S. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In: *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 117–124).

Pardos, Z. A., Gowda, S. M., Baker, R., & Heffernan, N. T. (2012). The sum is greater than the parts: Ensembling models of student knowledge in educational software. *ACM SIGKDD Explorations Newsletter, 13*(2), 37–44.

Pavlik, P., Cen, H., & Koedinger, K. R. (2009) Performance factors analysis—A new alternative to knowledge tracing. In: *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 531–538).

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering, 21*(6), 759–772.

Rai, D., & Beck, J. (2011). Exploring user data from a game-like math tutor: A case study in causal modeling. In: *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 307–313).

Rau, A., & Scheines, R. (2012). Searching for variables and models to investigate mediators of learning from multiple representations. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 110–117).

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In: *Proceedings of the 13th International Conference on Artificial Intelligence in Education, Marina del Rey, CA* (pp. 203–210).

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135–146.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, 40*(6), 601–618.

Rus, V., Moldovan, C., Graesser, A., & Niraula, N. (2012). Automated discovery of speech act categories in educational games. In: *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 25–32).

San Pedro, M., Baker, R., Bowers, A., & Heffernan, N. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 177–184).

Sao Pedro, M., Baker, R., Montalvo, O., Nakama, A., & Gobert, J. D. (2010). Using text replay tagging to produce detectors of systematic experimentation behavior patterns. In: *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 181–190).

Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1998). The TETRAD project: Constraint based aids to causal model specification. *Multivariate Behavioral Research, 33*(1), 65–117.

Scheuer, O., & McLaren, B. M. (2011). *Educational data mining. The encyclopedia of the sciences of learning*. New York: Springer.

Schreurs, B., Teplovs, C., Ferguson, R., De Laat, M., & Buckingham Shum, S. (2013). Visualizing social learning ties by type and topic: Rationale and concept demonstrator. In: *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 33–37).

Shanabrook, D. H., Cooper, D. G., Woolf, B. P., & Arroyo, I. (2010). Identifying high-level student behavior using sequence-based motif discovery. In: *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 191–200).

Shute, V. J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Interaction, 5*(1), 1–44.

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254).

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. New York: MIT Press.

Srikant, R., & Agrawal, R. (1996). *Mining sequential patterns: Generalizations and performance improvements*. Heidelberg, Germany: Springer.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics, 31*(6), 2013–2035.

Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 64–74).

Tatsuoka, K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327–359). London: Routledge.

Thai-Nghe, N., Horvath, T., & Schmidt-Thieme, L. (2011). Context-Aware factorization for personalized student's task recommendation. In: *Proceedings of the International Workshop on Personalization Approaches in Learning Environments* (pp. 13–18).

Vuong, A., Nixon, T., & Towle, B. (2011). A method for finding prerequisites within a curriculum. In: *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 211–216).

Part II
Learning Analytics for Learning Communities

Chapter 5
Analytics Through an Institutional Lens: Definition, Theory, Design, and Impact

Matthew D. Pistilli, James E. Willis III, and John P. Campbell

Only about a decade after initial projects emerged, learning analytics have already had a profound effect on how institutions view the combination and use of multiple data sets and their subsequent analysis on learners and instructors. The success or failure of these institutional projects lay in the confluence of the data that is available, the extent to which analysis is conducted on that data, and the ensuing action taken on the results. Some higher education institutions have begun using the power of analytics to affect positive outcomes in critical areas like learning, pedagogy, student retention, and institutional decision-making (Long and Siemens 2011). But what does analytics mean to an institution and how might an institution implement some form of analytics? Further, what can institutions expect from the successful implementation of analytics?

This chapter begins with a broad description of analytics from an institutional perspective, including the foundational theory from which institutions can build. From there, what happens to an institution's learning environment when a successful implementation occurs, as well as how various learning communities are affected, will be examined. Finally, a discussion directly of how institutions may be changed as a result of data-driven models to enhance levels of success will be presented.

5.1 Defining Analytics from an Institutional Perspective

Bichsel (2012) defines analytics as "the use of data, statistical analysis, and explanatory and predictive models to gain insights and act on complex issues" (p. 6). Campbell et al. (2007) liken this joining of institutional data sets to a marriage—one that allows for the institution to discern patterns of student behaviors, traits, or outcomes.

M.D. Pistilli (✉) • J.E. Willis III
Purdue University, West Lafayette, Indiana, USA
e-mail: mdpistilli@gmail.com

J.P. Campbell
West Virginia University, Morgantown, West Virginia, USA

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_5, © Springer Science+Business Media New York 2014

The identification of patterns provides an institution with the opportunity to provide targeted actions (interventions) to improve student learning. As these scholars and others have noted, a necessary component in learning analytics is taking action; institutions cannot simply collect and report on data. Institutional investments must go beyond performing data analysis and actually take specific actions to enhance student success and institutional achievement. These actions can include emailing or sending letters to students with specific steps that can be taken to enhance course performance (Pistilli and Arnold 2010; Arnold and Pistilli 2012), providing tips to students as to what it took other students with similar academic backgrounds to be successful in a course (Bramucci and Gaston 2012; McKay et al. 2012), or developing programs designed to mitigate various behaviors (Frankfort et al. 2012; Taylor and McAleese 2012).

The concept of using data to take action, to address something, or to facilitate decision-making processes is not new. Corporations have long used data on consumers and their habits to determine marketing strategies, directions for product development, and predicting sales based on current buying habits. They call this practice business analytics, which is defined as "the practice of iterative, methodological exploration of an organization's data with emphasis on statistical analysis. [It] is used by companies committed to data-driven decision making" (Rouse 2010).

Using data to drive decision-making processes is not new to the higher education sphere, either. Colleges and universities have begun using data to better understand and begin to address student success, retention and graduation rates, course offerings, financial decisions, hiring and staffing needs, or admissions models of admits, yield, and matriculation. Furthermore, the use of analytics allows institutions "to test … assumptions [regarding theoretical, practice-based and/or evidence-based examples of sound educational design] with actual student interaction data in lieu of self-report measures," making for much more compelling arguments for continuing a process once one has been shown to be effective (Lockyer and Dawson 2011, p. 155). Analytics should take these tests and the broad use of data one step further. Analytics moves an institution, and the realm of education, from simply understanding various data points and their intersections, to using them to create actionable intelligence—and then taking action on that intelligence as a means of positively affecting one or more behaviors or outcomes.

Early systems like Purdue University's *Signals* demonstrate how an institution can use analytics to create actionable intelligence that give students tools for real-time change (Pistilli and Arnold 2010). The key term here is "actionable" because theoretical knowledge of analytics, while perhaps valuable administratively, does little to influence student behaviors; rather, specific direction to help struggling students is what really matters (Cooper 2012). Long and Siemens (2011) identify key ideas driving the "value" of learning analytics, noting that their "role in guiding reform activities in higher education, and … how they can assist educators in improving teaching and learning" (p. 38).

Although learning analytics is fairly new, higher education institutions do not have to explore the use of analytics without a guide. Educational research has provided decades of studies waiting to be transformed into practice. For example, Google Scholar provides over 274,000 articles on "prompt feedback" and analytics may

become the basis by which institutions can bring the feedback literature to scale. Whether looking at foundational research from Astin and Tinto on student success or focusing on Chickering and Gamson's seven principles, analytics provides the opportunity for scaling decades of educational research into daily practice.

5.2 Theoretical Basis for Institutions Implementing Analytics

Institutions should not arbitrarily approach analytics to change student behaviors, activities, or outcomes, as these kinds of efforts are already overwhelming to some institutions based on organizational and technical challenges alone. Given this, institutions should consult existing education theory and research to determine where best to apply resources. For many institutions, use of analytics to improve retention remains the key focus. While retention of students is a necessary thing to examine, it is an outcome measure. Effort should be put into improving student success, however success is defined by an institution, in such a way that a more successful student is more likely to be retained (Tinto 2012). To guide institutional efforts towards student success and, ultimately, retention, we turn to three guiding theories: Tinto's (1975, 1993) theory of student departure, Astin's (1984, 1993, 1996) theory of student involvement, and Chickering and Gamson's (1987) principles for good practice in undergraduate education.

Tinto's theory of student departure is the most widely cited theory in retention circles, and according to Braxton (1999), has reached "near paradigmatic stature" for those in the field of higher education (p. 93). In *Leaving College*, Tinto (1993) proposed that institutions need to meet three main conditions in order to achieve student persistence. First, students need to have access to retention programs that put their welfare above the institution's goals. While many institutions provide retention programs, analytics may provide a basis for better understanding who is attending such retention programs and how an institution may encourage all students to take advantage of the existing programming.

Second, retention programs should not just focus on a particular population (e.g., minority students, low-income students, athletes), but, instead, need to be available to all students from all walks of life. An institution's use of learning analytics may provide an opportunity to reexamine the use of student success programs. Rather than focus on particular categories of students, analytics may allow an institution to identify particular behaviors that change over time—providing a new dynamic learning environment that progressively evolves rather than remaining fixed on a particular group.

Third, retention programming must work to provide a degree of student integration within an institution to be successful. In his theory of student departure, Tinto (1993) notes that it is critical that students become formally and informally integrated to both their academic and social lives while in college. Formal academic integration occurs when a student interacts with a professor in class, visits a professor

to discuss class concepts during that professor's office hours, or attends a tutoring session or resource room to gain a better understanding of topics covered in class. Informal academic integration can be seen through students' interactions about course content with one another outside of class, or their attendance at a voluntary field trip that further explores the topics discussed during normal course time. Formal social integration is a function of students pledging fraternities or sororities, joining clubs, or participating in student government. Informal social integration revolves around students simply interacting with one another, such as playing video games together in a residence hall or playing a pick-up game of basketball. Tinto (1993) argued that the likelihood of persistence for students is increased if they have positive experiences with all four of these types of integrations.

For an institution, the formal and informal social integration becomes the most promising and challenging aspect for analytics. The potential data sources from these activities could provide a new insight to students integration into campus—however the collection and analysis of such data is probably beyond the capability of most institutions. Metrics that would indicate the achievement of a level of integration would need to be built into the interventions employed by an institution so that appropriate data could be collected to show the efficacy of the effort. It is possible that proxies may need to be developed by an institution if the exact data that would indicate an outcome is difficult, if not impossible, to obtain.

Tinto's theory provides a means of understanding the principles behind a student deciding to remain enrolled at an institution or to seek to be enrolled elsewhere, if at all. It should be noted that the decision to stay or go is first predicated on the extent to which students have been successful in their coursework. An unsuccessful student is likely to leave an institution, possibly of their own volition or more likely as a result of being academically dismissed. Analytics, then, is a process that can influence students' behaviors to help them be more successful, thus leading to their retention on campus. To this end, then, Tinto's model can be used by institutions as a roadmap for analytics.

The results of an institutional project based on Tinto's work could come in the form of encouraging a student struggling in one or more courses to actually visit the professor or available help rooms. Students who haven't joined a student organization might be encouraged to do so. Professional advisors and student resident assistants might be employed to outreach to specific students. In short, by looking at the integration scenarios and decision-making points, collecting data, and executing analytics, an institution can directly apply the "actionable intelligence" from the analytics to ensure that students are receiving the feedback needed, are connected to the institution, and are on track to graduate from that university in a reasonable amount of time.

Astin's theory of involvement (1984, 1993, 1996) noted that the more involved students are with certain aspects of their collegiate lives, the more likely they are to succeed. This notion was incorporated in the Input-Environment-Outcome (IEO)

college impact model (Astin 1993). Greatly simplified, outcomes, or characteristics of students once they have experienced collegiate life, are thought to be based on the initial characteristics students bring with them to college (input; e.g., gender, ethnicity, socioeconomic status) and are impacted by the collegiate experience as a whole (environment; e.g., going to classes, seeking academic help, participating in clubs/organizations). With regard to inputs, Astin (1993) identified 146 characteristics in several different groupings, including demographic, past academic achievement, previous experiences, and self-perception. Institutions may look at these characteristics as potential data elements for their analytics efforts.

Outcomes range from very concrete, easy to measure things such as level of academic achievement, retention from one year of study to the next, and persistence to graduation. More abstractly, there are skills, pieces of knowledge, and specific behaviors that are also developed. These tend to be the things that most institutions want to develop in students in some form. In the end, however, an institution has no impact on the inputs students bring with them to college, and only have the ability to potentially influence the achievement or development of various outcomes.

However, an institution can have a direct effect on student outcomes—and that is through the manipulation and alteration of the environment provided to the students. The environment consists of factors that are directly related to students' experiences while in college, many of which are influenced, if not directly offered, by an institution's administration. Astin (1993) identified 192 variables across eight classifications of characteristics that are associated with environment: institutional (e.g., Carnegie classification, size); peer group (e.g., socioeconomic status, values instilled, attitudes portrayed); faculty (e.g., teaching methods, interaction opportunities); curriculum (e.g., existence of a core set of courses, course requirements, delivery method of courses); financial aid (e.g., types of aid provided, amount awarded); major field choice; place of residence (e.g., on/off campus, fraternity/sorority housing); and student involvement in his/her education (e.g., hours spent studying, number of courses taken in a specific field).

Astin (1996) also concluded that there were three degrees of involvement for students at a given college or university. The first, involvement with academics, involves looking at the amount and quality of time spent on homework, working on projects, or engaging in other course-oriented activities. These activities are often the focus of many current institutional analytics projects. The second surrounds students' involvement with faculty within their courses and outside of the classroom. From an analytics perspective, these activities are often difficult to track as they usually require faculty compliance in noting interactions or attendance or some other form of recording of the students' involvement. The third degree, involvement, is the extent to which students interact with peer groups during college. These are fostered out of the collegiate environment, but are influenced both by the inputs students bring with them as well as their individual desired outcomes. These factors are important to keep in mind, because their interplay with the environment will have distinct effects on the extent to which the application of analytics is successful.

Where Tinto provides some insight into some of the behaviors that can be influenced as well as key decision-making points for a student, Astin is more concerned with altering the environment in which students exist. As a result, the input-environment-output model integrates nicely with the application of analytics in that it provides the ability to achieve a desirable outcome (success in college, increased learning, progression towards and awarding of a degree, etc.) by taking any number of inputs from a multitude of experiences and actions, and providing suggestions for behaviors and environmental interactions that would allow students to interact with one another, with faculty, and with the university in a positive and supportive manner. Astin's research becomes a foundation from which to begin an analytics project and allows for the analysis of many different pieces of data in coordination with each other to find meaningful relationships and areas to address. Using Astin's model, a framework has been created that provides variables based on, but different from, Astin's work to examine and behaviors to influence in an effort to achieve a desired outcome.

5.3 Institutional Analytics by Design

Tinto and Astin each provide a theoretical basis for student success and retention, but where does an institution turn to begin designing an analytics project?

The goal for an institution is to design a learning environment that leverages the institution's data, the seamless integration of pedagogy, and the actions required to have an impact on student success and retention. The ideal interaction for these three aspects has each informing the other to create the optimal mix for a given institution. For example, a given pedagogy will direct the kind of action(s) to be taken, as well the analysis of certain data points. Both specific actions and pedagogies will drive the kinds of data to be collected. The extent to which data *can* be gathered will certainly influence actions taken, and also may influence the extent to which a given pedagogy is used by an instructor. In the end, however, individual institutions must determine what kinds of interactions are the most meaningful given their unique needs and priorities (Fig. 5.1).

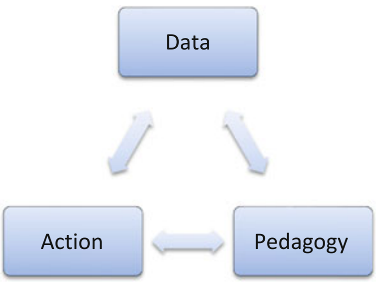

Fig. 5.1 The interactions that are needed in a learning environment to positively affect student success and retention. Each aspect informs the other to provide an institution with insight into what data to collect, analyze, and disseminate; what actions to influence or perform; and, which pedagogical practices to employ or examine

5.3.1 Designing for Data

For analytics to be successful, an institution must place an initial emphasis on collecting, organizing, and analyzing data that is meaningful, useful, and obtainable. As Campbell and Oblinger (2007) note, “data is the foundation of all analytics efforts” (p. 3); absent a strong foundation of good data, any analytics effort will likely fail. Therefore, an institution should place a considerable effort on determining what data is available—or could be obtained—that can provide meaningful insights into achieving the intended goal of the effort. Once institutions have established a practice of using analytics, attention may be focused on collecting new data identified by Tinto, Astin, or other researchers. Chickering and Gamson (1987) provide a foundation for what data might be meaningful through their analysis of 50 years of education research. The authors proposed seven principles of good practice for undergraduate education. These principles include developing educational environments that:

1. Encourage contact between students and faculty
2. Develop reciprocity and cooperation among students
3. Encourage active learning
4. Give prompt feedback
5. Emphasize time on task
6. Communicate high expectations
7. Respect diverse talents and ways of learning (1987, p. 3)

For example, if an institution may wish to focus on the contact between students and faculty, the data might include the number of exchanges in emails, discussion postings, or participation in office hours.

While one could imagine a large number of potential data sets, it is important to balance the practicality of collecting said data with the number of students who will actually have their data recorded. To that end, institutions should not go out and create new data for the purposes of implementing analytics. Rather, institutions should strive to utilize ambient data—data that already exists or is collected as a matter of course—that may be used as a proxy to measure or represent student success. Institutions can identify potential data sets by utilizing the work of Tinto, Astin, Chickering, and Gamson—then test those data elements through statistical analysis. For example, while a computer or algorithm can easily determine if an email exchange was made between a faculty member and a student, the system is likely not able to determine the content of the exchange—thus not able to determine if the message was to provide prompt feedback, communicate expectations, or clarify the purpose of the assignment—or something different altogether.

Taking this approach of using ambient data is reinforced by Macfadyen and Dawson, who write that learning analytics should be derived from student data points that are “readily accessible, scalable, and non-intrusive, and provide sound lead indicators of eventual student achievement or failure” (2010, p. 598). Data sets may include interaction data derived from learning management system (LMS)

logs, demographic information obtained through the admissions process, or past academic performance garnered as a result of students simply being enrolled at an institution, to name a few examples.

While self-report data can be highly valuable, researchers are dependent on high degrees of compliance with regard to completion of surveys or reporting of behaviors in order to make conclusions that can lead to action being taken. This is not to say that self-report data should be ignored altogether; these data are extraordinarily valuable in their own right. However, the collection and maintenance of self-report data is usually not sustainable. Self-report data are useful for confirming output from an algorithm using existing data to ensure that the targeted students are being identified correctly. The use of self-report data can also be influenced by the student's recognition of how the data may be used.

The paradox contained in the wealth of information contained in self-report data and the challenges associated with collecting and analyzing that data does not go unnoticed. On one hand, the collection of this kind of data comes down to the feasibility and sustainability associated with its gathering, analysis, and subsequent use. On the other hand, by collecting self-report data, an institution is making students aware that their data is being collected. This process, then, could lead to students automatically changing their behaviors—but could also result in students ceasing to provide information at all. The process of collecting any data, be it ambient or otherwise, should be focused on being both scalable beyond a pilot collection at one or two events *and* sustainable so that the data collection effort isn't abandoned because of efforts or challenges associated with it.

If institutions choose to utilize self-report data, it is important for them to consider whether the return on the investment associated with these actions is worth the information that it may potentially yield. If not, institutions may wish to look for other means of determining student behavior or action, or develop variables that approximate the desired outcome. Ultimately, an institution's decision to collect non-ambient data should be done so in the ways they feel are best for their individual purposes.

Campbell and Oblinger (2007) suggest that the manner in which data is stored and the length of time it is retained should be examined when determining what data will be used and the extent to which historical data will be analyzed to make real-time predictions. In addition, they also indicate that the granularity of data be considered as well. Defined as "a balance between what the system provides, what questions the institution is attempting to answer, and the storage requirements of the data," granularity refers to just how finite a set of data needs to be created (Campbell and Oblinger 2007, p. 5).

Differentiation needs to be made for current analytics projects and potential analytics projects. A handful of data points may accurately predict a given phenomenon equally as well as dozens of data points for the same event. To this end, then, an institution may strive for parsimony in its data sources for two reasons. First, institutions should create as sustainable and easy a process as possible for obtaining and analyzing data, and the fewer the number of points that need to be collected, the more likely it is that this end will be achieved. Second, in striving for something that

can be easily interpreted by faculty and students, the smaller the number of pieces of information utilized, the easier the institutions will be able to pinpoint exact areas that can be altered or addressed by either party. Finally, storing data for potential future use should be a separate function from current analytics projects.

5.3.2 Designing for Pedagogy

With the nature of today's electronic tools, designing for pedagogy is an interaction between two components—the process and the environment. From a process perspective, the success of an institution's analytics depends on the quality and quantity of the data that is utilized.

Institutions can design for analytics projects by integrating Chickering and Gamson's (1987) model as a basis for pedagogical design—ensuring that course design and analytics work synergistically. For example, the more that an instructor meaningfully includes tools that encourage active learning, provides prompt feedback, and increases communication between students that not only improves learning, but also provides a meaningful data set for analytics. Using a tool that allows faculty to email students regarding their performance and activities in which they can engage to enhance or maintain their current grades meets the first and fourth principles directly, and potentially the fifth, sixth, and seventh, depending on the content of the message (see Sect. 5.3.1 in this chapter for a numbered list of practices). By using analytics to determine teams or groups in classes, the course design fulfills the second principle. The third principle, use of active learning techniques, might have an analytic solution that provides students with tailored practice exams, readings, or exercises that meet their specific needs or deficiencies. The broad point here is that the application of analytics can provide an institution with a means of not only achieving certain metrics or learning gains, but also creating an environment that is best suited and supportive of students and faculty alike.

When institutions use analytics as a process, they provide a means for evaluation and enhancement of course design through the alteration and enhancement of pedagogical practice. Analytics processes that are implemented well should identify courses where students continually struggle, and, further, specific instructors who may need to alter the manner in which they deliver material. In turn, pedagogy can be altered and enriched; by offering feedback to students where none was offered in the past, the instructor can shift the focus of instruction to students' needs in an effort to help them improve. Over time, as students receive more and better feedback, and as faculty recognize areas for improvement in their own modes of instruction, the quality of these courses should improve. Instructors will have a better understanding of how material is being received, can address specific topics, and potentially move more quickly through the course. Students, then, are learning more and are better prepared for subsequent courses. (For insight into how the application of analytics can improve student success and other outcomes, see Arnold 2010; Arnold and Pistilli 2012; Baepler and Murdoch 2010; Essa and Ayad 2012; Ferguson

2012; Oblinger 2012; Salas and Alexander 2008; Taylor and McAleese 2012; or Tinto 2012.) The secondary courses are then able to cover more material, since students will not need to be remediated on core concepts. In all, the institutional application of analytics can result in a major shift for colleges and universities with regard to the culture fostered around undergraduate learning.

As mentioned above, the use of analytics can alter the structure of the classroom and, indeed, the institution itself. However, as noted by Bonfiglio et al. (2006) the transition to a student-focused environment from one that was more suited for the instructor takes a great deal of institutional analysis, with particular emphasis placed on the extent to which learning is taking place. This can be done in many ways, but is increasingly seen within the realms of computer-supported collaborative learning, social learning, and distance learning.

From an environmental perspective, the institutional culture will have a significant impact on how an individual faculty member designs courses with analytics in mind. The availability of data, willingness to allow faculty to experiment, and trust between data "holders" and data "users" will all have an effect on the extent to which faculty can utilize data to change their pedagogies—and the extent to which an institution is willing to exert effort to collect additional data to inform pedagogy. The interaction between those that collect, hold, and utilize the data will set the tone as to how analytics will be fostered—or abandoned—at an institution. An institutional culture that controls data tightly or seeks to use data as an evaluation hammer will limit the use of analytics to improve pedagogy. Conversely, an institutional culture that allows more liberal use of data may foster new pedagogical models based on analytics.

5.3.3 Designing for Feedback

As Lockyer and Dawson (2012) note, though, analytics and resulting actions are usually done "retrospectively—and often on an *ad hoc* basis" (p. 14). While Lockyer and Dawson's reflective view is both effective and necessary, Chickering and Gamson (1987) suggest providing feedback at a point in a term where students have the opportunity to alter their academic behaviors, thus giving students the biggest chance of success. Lockyer and Dawson (2012) also posit that the application of learning analytics needs to be in synch with what instructors are planning to do inside of their courses. Data needs to be appropriately chosen for analysis, and provided in such a manner that it is easily interpreted by and broadly useful to the instructor – and that is ultimately meaningful to the student. Utilizing Chickering and Gamson's framework, designing courses with feedback as an integrated piece would look for solutions that not only provide for prompt feedback, but also seek ways to encourage communication between the faculty and student, continue to set high expectations, and direct students towards more active learning approaches.

Schunk (1985) notes that students gain self-efficacy and knowledge from having the opportunity to work at a task and receive both positive and negative feedback.

Furthermore, he indicates that giving students training in various strategies that could be used to address a problem, and then allowing them to determine which one is most effective, provides opportunities for students to better understand the required task. Finally, feedback is necessary for students to fully understand their performance, their shortcomings, and their successes. All of these activities, particularly if they're done in an online environment, create large quantities of useful data that can be extracted and mined for patterns of performance, areas of weakness or strength, or prediction of future grades or outcomes. Another avenue for creating these data points would be through practice tests and quizzes offered through various technologies where data can be readily harvested. The point is not to create busy work, but, rather, to create data that can be analyzed and acted upon in a timely manner that will allow students to focus their studying efforts on areas of deficiency. When coupled with precise feedback and specific things to address, students will have the best opportunity to improve their performance.

5.4 Impact of Analytics

There are distinct differences between the types of communities that can be formed via brick and mortar classrooms or virtual learning environments, and the application of analytics can be useful in either one. In the end, however, the focus should be less on the mode of learning and content delivery, and more on the learning itself. This becomes readily apparent as the realized and potential effects that learning analytics has on learners, instructors, and institutions are examined.

5.4.1 The Impact of an Institution's Use of Analytics on Learners

If the driving force behind an institution's decision to delve into learning analytics is increased learning, then the learners must be the direct beneficiaries of such efforts. Assessing *how* learning is affected is best done thematically; though this list is not exhaustive, the current trends of learning analytics typically fall within examining predictive power, providing change opportunities with near live-time feedback, pulling together big data in ways that were impossible previously, and identifying knowledge gaps in learners.

One of the most promising aspects of learning analytics is the power of prediction based on historical and current data points (Elias 2011). Student predictions are not static because if the model is robust enough, constantly changing data will sharpen the prediction "and then feed those results back in order to improve the predictions over time … as it relates to teaching and learning practices" (Elias 2011, p. 5). Though it may not be prudent to describe the process as somewhat parasitic, the metaphor works: if modeled correctly, data points help build and then clarify the

algorithm to statistically predict a learning outcome. Put into real-world terms, this means that seemingly disparate data on a given student may be brought together meaningfully to provide actionable information for the student.

Macfadyen and Dawson (2010) demonstrated this point by "providing data from an international research project investigating which student online activities accurately predict academic achievement" (p. 588). They found that providing instructors with a "dashboard-like interface that incorporates predictive models and network visualization tools" was able to provide statistically significant predictive student outcomes, such as increased engagement with peers and course material and "higher overall final grades" (p. 597). There is a sharp distinction between prediction and actuality; as the power of predictive learning analytics increases over time, all parties involved in the learning process must heed the differentiation between what a statistical model can predict as an outcome and how students' *choice of action* once presented with the data affects the actual outcome.

Learners are ultimately responsible for how they take analytical predictions and create an opportunity for change. What drives this change mechanism is data mining, which, in this context, "discover[s] potential student groups with similar characteristics and reactions to a particular strategy" (Romero et al. 2008, p. 1). Amalgamating and assessing large data sets of student demography, history, performance, and numerous other data points have the potential to provide near real-time feedback. At the granular level, data mining can provide expeditious information that, among other things, helps "to identify learners with low motivation and find remedial actions to lower drop-out rates" (Romero et al. 2008, p. 1). Fostering the shift from using large data sets to real, quantifiable information that can nudge a student to make a real change are the algorithms that drive institutions' LMSs; the information that institutions use for their formulae is already at their fingertips. The task in learning analytics is to take information from the LMS which functions in near real-time and convert it into a compelling interface to encourage student change; the process of converting LMS data into a student nudge has the potential "to offer students a truly flexible and rich learning experience" (Phillips et al. 2011, p. 1005).

The use of algorithms generated from large data sets for the purposes of learning analytics presents something of a contemporary phenomenon: the convergence of big data for actionable intelligence. The consequences of using this data, which can persuade students to make meaningful change in their courses of study, have directly measurable outcomes, including "retention, graduation in a timely manner, and preparation for the workforce and citizenship" (Chacon et al. 2012, p. 7). Making meaning out of large data sets, especially as they apply to student learning, is no easy task; this requires being able to think outside of LMS-generated data (Pardo and Kloos 2011). As Pardo and Kloos (2011) highlight, the "interaction" of data is what really matters because it is where the "learning experience" is measured (p. 163).

Learning analytics, especially as it is used as a measure of the "learning experience," helps identify knowledge gaps in students. A student's performance on an

assessment may be indicative of learning gaps of specific information; when coupled with performance on multiple assessments, patterns may emerge that show that student's learning deficiencies in an entire area. The difficulties of measuring these patterns, though, are akin to measuring students' learning with perceptions. As Phillips et al. (2011) point out, learning measured with perceptions "rarely indicate the causality of effects" (p. 998). Measuring patterns in student performance may not necessarily highlight the causality of the effects, therefore indicating actionable change may be a product of correlation, not causation.

The impact of learning analytics on students may also have some poignant challenges including not knowing what students will do with learning analytic data, suffering from self-fulfilling prophecies, and the effects of relinquishing data. There is an element of the unknown insofar as analytics are not necessarily the actual outcome of a given set of variables. Perhaps the more potent idea is that student-centered data, when constructed in a rigorous and ethically-sound model, can empower students to make responsible changes to enact more favorable outcomes. Limitations like the innate intrusiveness of gathering and using data present unique challenges; the priorities of helping individual students must be weighed against the potential problems of using personally-identifiable variables.

Predictive analytics works in a way that lives up to its name: using yesterday's and today's data to predict tomorrow's outcome. With learning analytics, institutions cannot know what students will do with information regarding performance, suggestions for improvement, or predictions of outcomes. Mattingly et al. (2012) draw this out explicitly in terms of "distance education" when they note "the lack of knowledge about the ways that students interact with learning materials" (p. 238).

Highlighting the dearth of information available on how students interact with learning materials, the parallel implication can be examined in terms of predicting how a student will use learning analytics. For example, Pardo and Kloos (2011) suggest today's students "rely less on the functionality offered by the LMS and use more applications that are freely available on the net" (p. 163). There is an element of Occam's razor here: students tend to use whatever resource is easiest to obtain and most expeditious in providing a result. The problem with learning analytics in this environment, then, is measuring how deviating from a set of metrics (often configured in an LMS) changes the model or the predictive power within a set of probabilities.

With any analytical process that relies on predictive power, there is a risk of impacting students for the worse via a potential self-fulfilling prophecy (Merton 1948). In other words, if students are given information that their prior performance on assessments indicates a real jeopardy of failing a course, they may simply give up and accept failure as a predetermined outcome (McKown et al. 2010). Even worse, if students interpret failure in one course as failure in general, they may drop out of school altogether. The challenge then lies in providing meaningful and actionable feedback to students in ways that will help them. Providing constructive information to the student helps overcome the communication barriers known to exist when encouraging a student to act on analytical knowledge (Tanes et al. 2011).

One of the consequences of using big data and learning analytics is the fact that the data being used has to be acquired. The interplay between large sets of group student activity (both academic and non-academic) coupled with granular student-level data is at once powerful, compelling, and worrisome. In this sense, "big data" can be that which is culled from numerous students or multiple data points from one student. What makes it "big" is not just the number of computable variables, but also the power of what that data might suggest through the power of statistical regression. This data can be obtained either through individuals giving it up (voluntarily or otherwise) or through the mining of existing data sets. (For a broader discussion of big data within the context of education and analytics, see also: Campbell et al. 2007; Ferguson 2012; Picciano 2012; or, Siemens and Long 2011.) Chacon et al. (2012) encourage "a clear policy framework in applying learning analytic tools and developing intervention strategies" because "institutional real-time monitoring of student progress … might be perceived as intrusive" (p. 7). This is especially true as social media posts, behavioral interactions, and other lifestyle data become more widely available and are shown to have efficacy with regard to predicting student success.

Like measuring learning, there is no definitive way to measure the effect of analytics on students. Learning analytics has provided institutions the ability to take existing data, compute predictions of students' performance, and intervene in near real-time to compel them to change their academic and/or social behaviors. Learning analytics seizes the power of data mining and applies data sets to quickly and efficiently identify how to help students improve their learning and performance on assessments. With this power of analysis, though, comes a bit of warning as to how such information may lead to self-fulfilling prophecy of failure as well as the relinquishment of what many may consider private data. One should consider the complexity of human learning; while data may provide some insights to student learning, most data today provides only a hint of current student progress. Presenting data as the one, authoritative source of student progress will cause unintended consequences – in the extreme, this may present itself as students or faculty giving up in the process. It is important to draw the potential connection between data sets that apply to cohorts of students as well as individual data: there is influence in metadata and individual data insofar as it affects the present and future of student success.

5.4.2 Impact on Instructors

Like learners, those who teach have gained immensely from learning analytics. In prior educational models, instructor feedback may have been measured in days and weeks; learning analytics has the ability to help instructors provide much quicker feedback. Additionally, learning analytics helps instructors reshape pedagogical practices because they have access to data-rich information on what works and what does not for a given class.

Perhaps one of the most difficult aspects of teaching is being able to assess if students are learning and, if so, to what extent. Greller and Drachsler (2012) provide a succinct summary of what learning analytics offers to instructors:

> Teachers can be provided with course monitoring systems that inform them about knowledge gaps of particular pupils and thus enable them to focus their attention on those pupils. They can also harvest emergent group models that can lead to shared understanding of domain topics or processes for better curriculum design and on-the-fly adaptations (p. 47).

The information flow from the analysis of data takes the same form as for students: information alone does not change anything, but can equip the instructor to make time-sensitive, meaningful changes that affect student outcomes. This requires what Elias (2011) calls "faculty buy-in," because it requires use of newer technologies, willingness to change long-held practices, and constructive criticism of methodologies of pedagogy. These elements are "paramount to the institution's ability to build and sustain a culture of evidence-based action" (p. 16). The power to shape student outcomes, however, is realized in how it affects individual students. Where many universities have lecture halls filled with several hundred students at a time, having the granular data to affect individual student outcomes is a direct benefit to an instructor's overall effectiveness.

When instructors use learning analytics to approach students with constructive criticism, effective communication is often what helps students succeed. If used properly, analytics enables instructors to "empower … students to monitor their coursework and take greater responsibility for their learning" (EDUCAUSE 2010, p. 2). This responsibility is important. The onus of action is still on the student, but the instructor is better prepared to guide the student with feedback in an amount of time necessary for changes in behavior. Feedback generated via learning analytics goes further than in previous educational models because it has the capacity to "go beyond reinforcement and [provide] an elaborate picture of where a learner stands in reference to others, certain criteria or their previous performance" (Tanes et al. 2011, p. 2415). Feedback in a timely manner provides a boon for instructors to affect student outcomes in real-time.

Having a proverbial yardstick by which to measure student learning is a key benefit of analytics, if for no other reason than it allows instructors the flexibility to alter pedagogical practices in a timeframe that positively affects student outcomes. If an indicator of pedagogical practice was once measured by end-of-course surveys (where changing teaching practices would only benefit future students), learning analytics provides feedback for instructors so that they can make quick alterations in teaching using real-time data. Using feedback as an aid in the teaching process does not necessarily yield a perfect or ideal pedagogy (van Harmelen and Workman 2012). This is due to a "multiplicity of uncontrolled variables," particularly because learning analytics is applicable to widely varying subject matters, teaching styles, and learning outcomes (van Harmelen and Workman 2012, p. 18). Learning analytics is not an elixir for ineffective teaching, nor does it reveal an ideal pedagogy; instead, it provides data-driven tools or suggestions to help instructors make changes that can be measured in terms of student outcomes.

Putting learning analytics into service for better teaching means thinking about the "learning environment" in terms of "flexible modalities for study" (Dawson et al. 2009, p. 185). Today, learning environments may be online, in traditional lecture or seminar with LMS integration, in massive open online courses (MOOCs), or in some hybrid combination of these modalities. What learning analytics does is help transform the various challenges of teaching in different modalities into actionable data that are primed for appropriate pedagogical changes. Dawson et al. (2009) discuss these recent transformations at length:

> Regardless of the overall didactic story telling [sic] and engagement prowess of individual presenters a shift to online necessitates a re-configuration in learning design and a conceptual shift in pedagogical practice. While this transition has been at ease for some educators—others have found the change rife with new complexities surrounding technology usability and integration in a context where communication cues and notions of student engagement are largely invisible (p. 190).

To develop the tools necessary to make pedagogical changes, current research indicates that models ought to be developed that consider "informative feedback [because it] is more effective in teaching desirable outcomes, and is perceived as more valuable by learners" (Tanes et al. 2011, p. 2415). As an example of an informative feedback system currently utilized, Purdue University's *Signals* operates in conjunction with the LMS to "provide both performance and outcome oriented feedback to students" (Tanes et al. 2011, p. 2415). This means that while instructors specify the parameters of performance for student feedback, that same data is used to assess whether the pedagogy being employed is effective for large groups of students. If an instructor has many "red lights" in the signal system (indicating students are in need of immediate performance alteration), then that instructor should realize that a pedagogical shift may be necessary in order for students to realize greater success. Pedagogical shifts that are made with "informed change" help instructors "provide evidence on which to form understanding and make informed (rather than instinctive) decisions" (van Harmelen and Workman 2012, p. 17). The benefit of learning analytics for instructors is the production and promulgation of hard data that allows for alterations in teaching method to be employed relatively quickly.

As a formalized system of research, learning analytics is relatively new. Dawson et al. (2009) acknowledge the "scarcity of resources available that can readily assist teachers in rapidly evaluating learning progress and behavior in order to better design learning activities to provide a more personalized and relevant learning environment" (p. 191). As more historical student performance data becomes available to researchers, better algorithms likely will be developed. Recent work in causal models have "identified links between certain measurable data attributes describing past student behavior and the performance of a student" but this, too, "is dependent on a body of historical data" (van Harmelen and Workman 2012, p. 17).

Preliminary results of measuring how learning analytics provide actionable data to instructors indicate that "student success was associated with instructional rather than motivational feedback, and type of rather than frequency of summative and formative feedback" (Tanes et al. 2011, p. 2420). The challenge of analytical data for these purposes is the sheer amount of "comprehensive" data needed to make the

case (Ali et al. 2012, p. 470). Multiple data points help bring out statistically significant patterns to refine algorithms relevant for feedback tools that impact pedagogy; the problem in the intermediary time, though, is the amount of data needed to compute such multivariate algorithms and a consensus on which data points are most useful.

As the ongoing work of learning analytics is used to help improve pedagogical practices, one of the important caveats to the research is ensuring that the data employed by instructors does not discourage students. Greller and Drachsler (2012) are quite emphatic on this point because they see that statistical modeling may box in "individual teachers or learners against a statistical norm" with the possible result of "strongly stifl[ing] innovation, individuality, creativity, and experimentation that are so important in driving learning and teaching developments…" (p. 47). Discouragement in students is an important measure when examining how algorithms compute pedagogical conclusions. Research in this area indicates that "positive feedback … generally emphasized performance, while negative feedback emphasized outcome" (Tanes et al. 2011, p. 2420). This research helps alter pedagogy because it can assist faculty on how to form appropriate messages to students based on specific parameters of success or failure and help institutions develop instruction. Appropriate wording in and construction of messages to students help mitigate the problems of student discouragement.

Learning analytics has impacted instructors in terms of forming pedagogical practices that are current with the types of learning environments seen in education today. Analytics provides instructors with tools to provide quick feedback as well as make rapid changes in pedagogical practice to affect student outcomes positively. As more data becomes available for study, more precise algorithms will be shaped to help instructors avoid discouraging students by providing accurate and actionable feedback.

5.4.3 Impact on Institutions

Where the effect of learning analytics for students and instructors is best seen as a microcosm wherein the outcome is measured with individuals and groups, the impact on institutions is a look at the macrocosm. Variables such as learning environments, student retention and graduation rates, and pedagogical effectiveness as measured by the achievement of positive student outcomes are all considered when measuring the impact of learning analytics on institutions. This effect is best examined by how learning analytics aids student retention and how institutions are able to refocus resources once specific areas are identified. These measures are examined through the prism of the potential difficulties of learning the "right" way to think about and use analyzed data, as well as the legal and ethical issues of using data for analytics (Johnson 2013; Willis et al. 2013).

Learning analytics has a direct, quantifiable effect on institutions that can be seen in how they affect student success and retention. EDUCAUSE (2010) identifies the

major goals of improving "student achievement, retention, and graduation rates and to demonstrate institutional accountability" through the major initiatives of "harness[ing] the power of analytics to develop student recruitment policies, adjust[ing] course catalog offerings, determin[ing] hiring needs, or mak[ing] financial decisions" (p. 1). Retention often becomes an institutional focus for using analytics because it can be a sound measure of how effective curriculum changes, effective recruitment, and institutional accountability are institution-wide. The difficulty with an institutional view of retention is that the data is reaped after a student has dropped out, revealing "gaping holes of delayed action and opportunities for intervention" (Long and Siemens 2011, p. 32).

As with retention, many current institutional learning analytics projects are frequently focused on binary predictions or results—whether the student will be retained, if the student is at risk, whether the student understands a particular learning outcome, and other similar propositions. This binary focus of learning analytics is more a result of the current sophistication of the data, models, and interventions than the long-term potential of learning analytics.

The task for assessing retention data quickly is best accomplished by using the vast resources of data that are already present institutionally; these data points can be utilized with analytics to "serve as a foundation for systemic change" (Long and Siemens 2011, p. 32). A definitive institutional strength to using learning analytics is the methodological process of performing "hypothesis-driven [analysis], using a particular dataset to solve a practical academic problem, such as increasing student retention levels" (Baepler and Murdoch 2010, p. 2). While retention is certainly an important measure of institutional effectiveness, it is also illustrative of the interconnectedness of other measures of change; retention may be the computed number, the quantitative measure, but it is the byproduct of how well an institution focuses and redirects its energies to ensure students are successful.

Identifying areas of focus for institutions often demands the use of "granular level" analysis that is possible through statistical modeling, prediction, and analytics (Greller and Drachsler 2012, p. 47). In an age of increasing accountability and tightening budgets, learning analytics provides data that "can support optimal use of both economic and pedagogical resources while offering a structure for improved educational outcomes" (EDUCAUSE 2010, p. 2). However, the art of designing an analytic system to account for system-wide variables in pointing out areas of concern means combining "principles of different computing areas (data and text mining, visual analytics and data visualization) with those of social sciences, pedagogy, and psychology" (Ali et al. 2012, p. 470). This is extremely difficult; it requires drawing "value from data in order to guide planning, interventions, and decision-making [as] an important and fundamental shift in how education systems function" (Siemens and Baker 2010, p. 253).

Granular data usage requires critical evaluation of those variables that are the most important for the model, the same data points that will lead to increased knowledge of where students are failing to gain "self-directedness, critical reflection, analytic skills, and evaluation skills" (Drachsler and Greller 2012, p. 129). The influx of variables require diligent analysis because one of the most acute problems

in learning analytics is not only how to interpret the data, but to understand why "even … the best evaluative algorithms can result in misclassifications and misleading patterns" (EDUCAUSE 2010, p. 2). Once the variables that most closely align with the desired model are determined, institutions may take actionable change to impact students and instructors.

The actionable changes that institutions may take are often dependent upon knowing the "right" way to think about and use the analyzed data. van Barneveld et al. (2012) echo the warning that "analytics is not a one-size-fits-all endeavor and that one has to consider that analytics is a goal-directed practice" (p. 2). Rushing to make decisions can have unintended effects on institutions; rather, as Kay and van Harmelen (2012) argue, "the analytics 'silver bullet' lies in the potential to derive and to act upon pre-emptive indicators, on 'actionable insights,' stepping beyond the long haul reactive measures" (p. 5). Careful consideration of the various factors that lead to interpreted data is vital to institutions because it is integrated with learning, teaching, and administrative variables. Consideration of these variables also must include legal and ethical impacts on institutions.

The legal impacts on institutions must be considered carefully because all data sets have a certain amount of liability attached to them. The primary beneficiaries of learning analytics are students and instructors, but their data may be used to directly affect institutions for the better (Drachsler and Greller 2012, p. 123). Comparable institutions may find it beneficial to share data for mutual gain, but anonymization of data is important to prevent litigation and ethical breaches of conduct (Drachsler and Greller 2012, p. 127). Ensuring anonymity within data is important to data handling and transfer. To this end, Powell and MacNeill (2012) specifically describe the current need for more data handlers and data scientists—individuals who are able to work "across teaching and administrative domains, to ensure that relevant actionable insights from data can be identified and acted upon in meaningful, measurable ways" (p. 3). Proper training in the handling, use, and anonymization of data is important to safeguard institutions from litigation.

Beyond legal aspects of learning analytics, ethical uses of data are also of important concern for institutions. Even though students are increasingly "born digital" and thus have "new expectations" of the ways their data will be used, institutions have an ethical obligation to protect the data and work within accepted research methodologies (Kay and van Harmelen 2012, p. 5). The widespread use of big data begs questions of "data ownership and openness, ethical use and dangers of abuse, and the demand for new key competences to interpret and act on learning analytics results" (Drachsler and Greller 2012, p. 120). The problem with ethical analysis is keeping pace with the speed of technological development; though there are ethical models that are directly applicable to data use, the constantly changing environment demands an openness of ethical questioning. Though there are no definitive answers to ethical questions being asked in learning analytics, perhaps what is most important is to remember that analytics "is much more about a personal and organizational perspective on using data for decision-making and action-planning and less about how it is processed in a computer; evaluating, planning and doing are human activities" (Cooper 2012, p. 7).

5.5 An Example of Application

This chapter ends with a brief overview of how one institution chose to contextualize data and act on it accordingly. As mentioned above, Purdue University developed a system called *Signals* in 2007, which was a direct offshoot of Campbell's (2007) dissertation investigating the extent to which data derived from the LMS could be used to predict student performance. The challenge was to identify the student at risk of doing poorly in a class using only data that was readily available; to wit, current course grades (e.g., test/assignment grades), past academic history information (e.g., standardized test scores and high school or current cumulative GPA), demographic descriptors, and data indicating the extent to which a student was interacting with the LMS.

As these various data points were examined, it became clear to the researchers that the integration of these data could be done so in a way that would provide an outlet for meaningful feedback to be provided to students by their instructors. The goal became to assist students and "help them understand both their current grades in their classes and what they can do to earn a higher grade while there is still time to act" (Pistilli and Arnold 2010, p. 23). Throughout the development, care was taken to focus on the behaviors that could be addressed. Specifically, the intention was to get students information about the specific actions they could take to positively affect their standing in a course. This information was to be written by the instructor and delivered via brief email messages and postings within the LMS site for the course (Arnold 2010). Pistilli and Arnold (2010) note that the posting in the LMS is accompanied by a color—green, yellow, or red—which serves as a primary indication of how a student is doing in a course. Clicking on the light revealed a message containing substantive suggestions as to what a student could do to increase performance in the class.

The novelty of *Signals*, as compared to other early warning systems, is that it took students' effort into account by measuring the extent to which a student was interacting with the LMS *and* comparing that interaction to the interaction levels of other students in the same class (Arnold 2010; Pistilli and Arnold 2010). What this resulted in was a means for the instructor to directly tell students exerting less effort that they were, in fact, not expending as much energy online as the rest of their peers, and by taking more initiative they might be able to improve their performance in the class. In addition, the system provides a dashboard to both instructors and academic advisors, allowing both parties to see a student's progression of signals over the course of a semester. This allows for either person to directly intervene with students when they see a disturbing or downward trend in student performance in an effort to help them be more successful (Pistilli et al. 2012). Further, intervention can be made in an early and timely manner so that students, as discussed earlier in this chapter, have the opportunity to change how they interact with the course and the institution.

5.6 Conclusion

This chapter has provided insight into what analytics can mean to an institution, how one might go about implementing analytics, and some of the expected outcomes of the application of analytics. Ultimately, however, analytics is more than just a tool. It is a framework for a process that can drive other institutional activities. The systematic collection and analysis of data that drives predictions for student success that can be acted upon and have the process refined over time is how an ideal implementation should be envisioned.

The manner in which this can effect and alter an institution is undeniable. A well-coordinated analytics implementation allows institutions to use existing data to identify and interpret trends that result in increased student success, retention, and graduation. It facilitates the refocusing of efforts and resources to identify, remediate, and enhance programs and services offered to students. It has forced institutions to carefully and deliberately consider how they contextualize data and, subsequently, act on it both legally and ethically. Furthermore, decisions have had to be made surrounding the types of data to use, how to collect it, and the extent to which an institution involves students in both the determination of use and collection of the data itself.

References

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education, 58*, 470–489.

Arnold, K. E. (2010). Signals: Applying academic analytics. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/signals-applying-academic-analytics

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). New York: ACM. Retrieved from http://dl.acm.org/ft_gateway.cfm?id=2330666. doi: 10.1145/2330601.2330666.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*, 297–308.

Astin, A. W. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.

Astin, A. W. (1996). Involvement in learning revisited: Lessons we have learned. *Journal of College Student Development, 37*, 123–134.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning, 1*(2), 1–9.

Bichsel, J. (2012, August). *Analytics in higher education: Benefits, barriers, progress and recommendations* (Research Report). Louisville, CO: EDUCAUSE. Retrieved from http://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf

Bonfiglio, R., Hanson, G. S., Fried, J., Roberts, G., & Skinner, J. (2006). Assessing internal environments. In R. P. Keeling (Ed.), *Learning reconsidered 2: Implementing a campus-wide focus on the student experience* (pp. 43–49). Champaign, IL: Human Kinetics. Retrieved from http://www.myacpa.org/pub/documents/LearningReconsidered2.pdf

Bramucci, R., & Gaston, J. (2012). Sherpa: Increasing student success with a recommendation engine. In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 82–83). Retrieved from http://delivery.acm.org/10.1145/2340000/2330625/p82-bramucci.pdf. doi: 10.1145/2330601.2330625.

Braxton, J. M. (1999). Theory elaboration and research and development: Toward a fuller understanding of college student retention. *Journal of College Student Retention, 1*(2), 93–97.

Campbell, J. P. (2007). *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study*. Unpublished doctoral dissertation. West Lafayette, IN: Purdue University.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 40–52. Retrieved from http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume42/AcademicAnalyticsANewToolforaN/161749

Campbell, J. P., & Oblinger, D. G. (2007, October). Academic analytics. Washington, DC: EDUCAUSE. Retrieved from http://www.educause.edu/ir/library/pdf/PUB6101.pdf

Chacon, F., Spicer, D., & Valbuena, A. (2012, April 10). Analytics in support of student retention and success. *Research Bulletin 3, EDUCAUSE Center for Applied Research* (pp. 1–9). Retrieved from http://www.educause.edu/library/resources/analytics-support-student-retention-and-success

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 39*(7), 3–7. Retrieved from http://www.eric.ed.gov/PDFS/ED282491.pdf

Cooper, A. (2012, November). What is analytics? Definition and essential characteristics. *JISC CETIS, 1*(5), 1–10. Retrieved from http://publications.cetis.ac.uk/2012/521

Dawson, S. P., Macfadyen, L., & Lockyer, L. (2009). Learning or performance: Predicting drivers of student motivation. In R. Atkinson & C. McBeath (Eds.), *Ascilite 2009: Same places, different spaces* (pp. 184–193). Auckland, NZ: Ascilite.

Drachsler, H. & Greller, W. (2012). The pulse of learning analytics: Understandings and expectations from the stakeholders. In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 120–129). New York: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2330634 doi: 10.1145/2330601.2330634.

EDUCAUSE. (2010, April). 7 things you should know about… analytics. *Creative Commons Attribution*. Retrieved from http://net.educause.edu/ir/library/pdf/ELI7059.pdf

Elias, T. (2011). Learning analytics: Definitions, processes and potential. *Creative Commons*. Retrieved from: http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Essa, A. & Ayad, H. (2012). Student success systems: Risk analytics and data visualization using ensembles of predictive models. In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 158–161). New York: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2330641. doi: 10.1145/2330601.2330641.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5), 304–317. doi:10.1504/IJTEL.2012.051816.

Frankfort, J., Salim, K., Carmean, C., & Haynie, T. (2012, July). Analytics, nudges and learner persistence. *EDUCAUSE Review Online*. Retrieved from: http://www.educause.edu/ero/article/analytics-nudges-and-learner-persistence

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology and Society, 15*(3), 42–57.

Johnson, J. A. (2013, May 8). Ethics of data mining and predictive analytics in higher education. *Association for Institutional Research Annual Forum, Long Beach, California, May 19-22, 2013*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2156058

Kay, D., & van Harmelen, M. (2012). Analytics for the whole institution: Balancing strategy and tactics. *JISC CETIS, 1*(2), 1–31. Retrieved from: http://publications.cetis.ac.uk/wpcontent/uploads/2012/11/Analytics-for-the-Whole-Institutionvol1-No2.pdf

Lockyer, L. & Dawson, S. (2011). Learning designs and learning analytics. *LAK '11: Proceedings from the 1st International Conference on Learning Analytics and Knowledge* (pp. 153–156). New York: ACM. doi: 10.1145/2090116.2090140.

Lockyer, L. & Dawson, S. (2012). Where learning analytics meets learning design (workshop summary). In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 14–15), New York: ACM. Retrieved from http://delivery.acm.org/10.1145/2340000/2330609/p14-lockyer.pdf. doi: 10.1145/2330601.2330609.

Long, P., & Siemens, G. (2011, September/October). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 31–40.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*, 588–599. doi:10.1016/j.compedu.2009.09.008.

Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal, 4*(3), 236–247.

McKay, T., Miller, K., & Tritz, J. (2012). What to do with actionable intelligence: E^2Coach as an intervention engine. In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 88–91). New York: ACM. Retrieved from http://delivery.acm.org/10.1145/2340000/2330627/p88-mckay.pdf. doi: 10.1145/2330601.2330627.

McKown, C., Gregory, A., & Weinstein, R. S. (2010). Expectations, stereotypes and self-fulfilling prophecies in classroom and school life. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 256–274). New York: Taylor & Rutledge.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review, 8*(2), 193–210. doi:10.2307/4609267.

Oblinger, D. (Ed.). (2012). *Game changers: Education and information technologies*. EDUCAUSE. Retrieved from http://net.educause.edu/ir/library/pdf/pub7203i.pdf

Pardo, A. & Kloos, C. D. (2011). Stepping out of the box. Towards analytics outside the learning management system. *LAK '11: Proceedings from the 1st International Conference on Learning Analytics and Knowledge* (pp. 163–167). New York: ACM. doi: 10.1145/2090116.2090142.

Phillips, R., Maor, D., Cumming-Potvin, W., Roberts, P., Herrington, J., Preston, G., et al. (2011). Learning analytics and study behaviour: A pilot study. *Proceedings Ascilite Hobart 2011: Full Paper, 997-1007*. Retrieved from: http://www.ascilite.org.au/conferences/hobart11/downloads/papers/Phillips-full.pdf

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20. Retrieved from http://www.eric.ed.gov/PDFS/EJ982669.pdf

Pistilli, M. D. & Arnold, K. E. (2010, July–August). Purdue signals: Mining real-time academic data to enhance student success. *About Campus, 15*(3), 22–24. doi: 10.1002/abc.20025.

Pistilli, M. D., Arnold, K. E., & Bethune, M. (2012, July). Signals: Using academic analytics to promote student success. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/signals-using-academic-analytics-promote-student-success

Powell, S. & MacNeill, S. (2012). Institutional readiness for analytics. *JISC CETIS 1*(8), 1–11. Retrieved from http://publications.cetis.ac.uk/2012/527

Romero, C., Ventura, S., Espejo, P. G., & Hervás, C. (2008). Data mining algorithms to classify students. *Educational Data Mining 2008: The 1st International Conference on Educational Data Mining, Montréal, Québec, Canada, June 20-21, 2008 Proceedings, 8-17*. Retrieved from http://www.educationaldatamining.org/EDM2008/uploads/proc/full%20proceedings.pdf#page=8

Rouse, M. (2010). Business analytics definition. Retrieved from http://searchbusinessanalytics.techtarget.com/definition/business-analytics-BA

Salas, G., & Alexander, J. S. (2008). Technology for institutional enrollment, communication, and student success. In R. Junco & D. M. Timm (Eds.), *Using emerging technologies to enhance*

student engagement, New Directions for Student Services (124) (pp. 103–116). San Francisco: Jossey-Bass. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/ss.298/pdf. doi: 10.1002/ss.298.

Schunk, D. H. (1985). Self-efficacy and classroom learning. *Psychology in the Schools, 22*(2), 208–223. doi:10.1002/1520-6807(198504)22:2<208::AID-PITS2310220215>3.0.CO;2-7.

Siemens, G., & Baker, R. S. J. D. (2010). Learning analytics and educational data mining: Towards communication and collaboration. In S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). New York: ACM. Retrieved from http://delivery.acm.org/10.1145/2340000/2330661/p252-siemens.pdf. doi: 10.1145/2330601.2330661.

Siemens, G. & Long, P. (2011, September/October). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30–32.

Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using *Signals* for appropriate feedback: Perceptions and practice. *Computers & Education, 57*(4), 2414–2422. doi:10.1016/j.compedu.2011.05.016.

Taylor, L., & McAleese, V. (2012, July). Beyond retention: Using targeted analytics to improve student success. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/beyond-retention-using-targeted-analytics-improve-student-success

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*, 89–125.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Tinto, V. (2012). Moving from theory to action: A model of institutional action for student success. In A. Seidman (Ed.), *College student retention: Formula for student success* (2nd ed., pp. 251–266). Lanham, MD: Rowman & Littlefield.

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative, 1*, 1–11. Retrieved from http://net.educause.edu/ir/library/pdf/ELI3026.pdf

van Harmelen, M., & Workman, D. (2012, November). Analytics for learning and teaching. *JISC CETIS, 1*(3), 1–40. Retrieved from http://publications.cetis.ac.uk/2012/516

Willis, J. E., III, Campbell, J. P., & Pistilli, M. D. (2013, May 6). Ethics, big data, and analytics: A model for application. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application

Chapter 6
A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering

Andrew E. Krumm, R. Joseph Waddington, Stephanie D. Teasley, and Steven Lonn

6.1 Introduction

As demonstrated throughout this volume, colleges, and universities are increasingly finding value in aggregating and analyzing once disparate sources of data, such as a student's admissions records, academic history, and use of campus information technologies—all under the rubric of "learning analytics" (LA) (see also, Campbell et al. 2007; Fritz 2011; Goldstein and Katz 2005). In this chapter, we describe a design-based research project that developed Student Explorer, an early warning system (EWS) for an undergraduate engineering advising program. This project was organized around identifying user needs, developing the necessary infrastructure for building Student Explorer, and identifying factors affecting advisors' decision-making related to the use of Student Explorer.

The EWS described in this chapter represents an application of LA that is gaining popularity across colleges and universities—the near real-time aggregation and analysis of data on students' use of information technologies, such as Learning Management Systems (LMSs), for the purposes of identifying students in need of academic support (e.g., Beck and Davidson 2001; Macfadyen and Dawson 2010; Morris et al. 2005). One of the many benefits of collecting and analyzing LMS data is that these systems are used by a majority of instructors and students on most

A.E. Krumm (✉)
Center for Technology in Learning, SRI International, Menlo Park, CA, USA
e-mail: andrew.krumm@sri.com

R.J. Waddington
Institute for Educational Initiatives, University of Notre Dame, Notre Dame, IN, USA

S.D. Teasley
School of Information, University of Michigan, Ann Arbor, MI, USA

USE Lab, University of Michigan, Ann Arbor, MI, USA

S. Lonn
USE Lab, University of Michigan, Ann Arbor, MI, USA

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*, DOI 10.1007/978-1-4614-3305-7_6, © Springer Science+Business Media New York 2014

campuses in the United States (Dahlstrom et al. 2011; Fritz 2011). While there is increasing interest in using LMS and other related sources of near real-time data, few researchers document the specific ways users make sense of and base decisions on data generated by these systems. Moreover, few researchers connect the ways interested parties, such as academic advisors, provide support strategies to students using data and analyses from LA-based systems (Johnson et al. 2011; Lonn et al. 2012).

Student Explorer aggregates data from an LMS used at a large research university and provides near real-time data from that system to academic advisors in a program called the STEM (Science, Technology, Engineering, and Mathematics) Academy. The STEM Academy is a holistic student development program aimed at increasing the academic success of students who have historically been underrepresented in STEM fields. The STEM Academy is modeled on the Meyerhoff Scholars Program at the University of Maryland-Baltimore County (Maton et al. 2000) and the Biology Scholars Program (BSP) at the University of California, Berkeley (Matsui et al. 2003). Student Explorer provided the program's three academic advisors with frequent updates on students' academic progress and streamlined the presentation of data to allow for the immediate identification of students in need of support.

Developing Student Explorer in collaboration with an effective support program, such as the STEM Academy, provided a unique opportunity to advance the field of LA by identifying the ways in which academic advisors can use an EWS to support their interactions with students. In the chapter that follows, we provide a short overview of LA research focusing on prior projects that used data generated by LMSs. In general, LA-based systems using LMS data can be characterized as either providing data directly to students or providing data to an intermediary who then interacts with students. These two characterizations imply different numbers of steps—and different affordances and constraints—related to the ways in which LA-based tools may be thought to affect desired outcomes. In what follows, we describe the development of Student Explorer through the lens of design-based research. Throughout our discussion of Student Explorer's development and use, we address the following overarching research question: "How did advisors use Student Explorer to inform their support activities with students?" We conclude this chapter by addressing future directions for LA research.

6.2 Using LMS Data

LMSs are ubiquitous in higher education (Dahlstrom et al. 2011). Depending upon the system, LMSs can track click-level data on a variety of user-actions (e.g., when a student accessed a course discussion, uploaded an assignment, or downloaded a course resource). Given LMS's ubiquity and the growing potential to track and store data on user-actions, LMS data is a ready source for LA research. While a goal of LA research is to collect and analyze evermore novel sources of data, a further goal

of LA research is to explore how individuals can best use these novel sources of data to support their decision-making (Duval 2011; May et al. 2011). In using data from LMSs, there are two distinct LA research agendas. One agenda involves aggregating data from online learning environments and providing these data directly to students; the other direction involves taking similar sources of data and providing them to an intermediary, such as a course instructor or academic advisor, who then acts on that data.

Researchers, such as Judy Kay and Susan Bull, demonstrate the benefits of displaying data directly to students through what they refer to as "open learner models" (see two special issues of *International Journal of Artificial Intelligence in Education* (2007), volume 17, issues 2 and 3). Another early example of a system that provides data directly to students is the Context-aware Activity Notification System (CANS). Within a distance education context, Goggins et al. (2010) found that students were able to use feedback provided by CANS to identify what their peers were doing, and what they, in turn, might need to do in order to catch up to their peers. Intelligent tutoring systems provide students with real-time directed scaffolding as they work to solve mathematics problems (Koedinger and Corbett 2006). E^2Coach at the University of Michigan provides tailored messages to students based on demographic *and* course performance data (McKay et al. 2012). These messages are designed to motivate students to take specific actions, such as allocating more time to prepare for exams. For direct-to-student LA-based systems, what data is presented to students and how it is presented appears to be an important area of research that is still very much under way.

Following the second approach, Black and Wiliam's (1998) seminal meta-analysis on formative assessment illustrates the potential for providing data to an intermediary, such as an instructor, to redirect students (for a recent review, see Hamilton et al. 2009). Work by Dawson et al. (2008), is one example of an LA-based system that provides LMS data to an intermediary—an instructor. They observed that when an instructor had data on students' use of an LMS, it allowed the instructor to identify students who were in need of support. Purdue University's Signals project is an example of an LA-based system that provides data to both instructors and students. This tool combines two other types of data to the LMS data: student demographic data and student grades (Campbell et al. 2007). These three data sources are formulated into a prediction model that assesses the likelihood of a student's academic failure. Instructors have the added ability to send messages to students based on a student's classification (i.e., red, yellow, or green) as designated by the system.

Across both direct-to-student and direct-to-intermediary LA-based systems, the user interacting with the system decides how to follow-up on the feedback. Some of these subsequent actions involve more "steps" than others, which adds complexity to the relationships between use of an LA-based system and intended outcomes. For direct-to-student systems, a necessary next action may be fairly clear, such as completing an extra problem set within an Intelligent Tutoring System, or more nuanced actions extending over time and contexts, such as engaging in more study time based on recommendations made by E^2Coach or Signals.

For direct-to-intermediary tools, such as the one developed in this project, the intermediary makes sense of the data and identifies potential recommendations for students. Because intermediaries are situated actors, their sensemaking draws on and is affected by multiple factors, such as their familiarity with the students, courses, and even individual instructors. This sensemaking represents an extra step, one that implicates an intermediary's capacity to make sense of, use, and connect data to specific actions for students. Dependent on the filter provided through the intermediary, the student then makes sense of the recommendation and chooses whether or not to act on that information.

With these multiple steps in mind and to support students' academic success, we designed Student Explorer to shorten the time frame from when academic advisors first become aware of a student in need of support and their intervention with that student. However, given the multiple steps that are implied by providing data to intermediaries, we are purposefully cautious in making claims about Student Explorer's impact. Moreover, in these first two phases of the research, we carefully identified where we could collect data (and from whom) in order to support our understanding of Student Explorer's role across complex interactions among advisors, students, and data. In what follows, we document the ways advisors engaged students differently as a result of having access to Student Explorer.

6.3 Data and Methods

Our research agenda is organized around principles of design-based research (Brown 1992; Collins 1992; Collins et al. 2004). Design-based research involves "a series of approaches, with the intent of producing new theories, artifacts, and practices that account for and potentially impact learning and teaching in naturalistic settings" (Barab and Squire 2004, p. 2). A distinguishing feature of design-based research is that the development of systems and theories is a collaborative effort among researchers and participants (Cobb et al. 2003). In our work, we collaborated with STEM Academy advisors on the development of Student Explorer and the ways in which Student Explorer could be used to support their existing work.

In the first two years of the STEM Academy, two years before Student Explorer was developed, advisors relied on students' self-reported grades that students brought to monthly meetings. According to advisors, the monthly meeting schedule did not provide frequent enough interactions between students and advisors. For example, once a student had failed an exam or assignment it was often too late to correct a student's academic trajectory. We were tasked with developing an EWS that STEM Academy advisors could use, at any point in the semester, to identify students in need of support. Student Explorer, therefore, was developed to increase the frequency with which an advisor contacts students.

6.3.1 Description of Student Explorer Data

Data used by Student Explorer are drawn primarily from the university's LMS. The LMS tracks interactions between a user and the system in the form of "events" and "tables." Events can include anything from the number of times a student accesses a course site to when a student downloads a specific course reading; tables are structured data, such as a course site's Gradebook (GB). Using data from different events and tables as well as various technological tools (e.g., R, Microsoft Excel, as well as database and web-authoring tools) we engaged in an iterative, collaborative design-based approach to develop Student Explorer.

We collected two generic sources of data from the LMS: grade data and log-in data. Grade data was collected by downloading each course site's GB and Assignments data. Log-in events were collected by querying the LMS's data repository and counting the number of times that a student accessed a specific course's site. Grades and log-in events were aggregated and translated into a variety of visualizations, including figures displaying students' grades compared to their peers over time along with lists of performances on individual GB entries. This information was updated and sent to advisors on a weekly basis. We developed a three-level classification scheme of Engage (red), Explore (yellow), and Encourage (green) that provided advisors with a simple depiction of the complex relationships between academic performance data, including longitudinal data and intra-course comparisons, and log-in events.

6.3.2 Methods

Design iterations occurred in two phases, corresponding with two academic semesters. Along with clarifying how LMS data could be integrated into visual displays and classification schemes, we also engaged in a variety of data mining activities between the two academic semesters. These data mining activities were used to identify patterns between a student's use of the LMS and his or her final course grade. We used functional data analysis techniques (Ramsay et al. 2009) to explore relationships between students' use of both the LMS in general and specific LMS tools with their final course grades across multiple engineering courses. We estimated smoothing splines for LMS tools used on a site across the 16 weeks of an academic semester using all students in the course and later subsets of students according to final course grade. This process allowed us to create smooth plots of LMS use over time and graphically explore LMS use across (1) final course grades, (2) time, and (3) courses. We also examined the first derivative of each of these plots, which yielded information about the week-to-week changes in course site log-ins and tool use.

To capture how Student Explorer was used by advisors and in their interactions with students, we conducted multiple individual and group interview sessions with

advisors. We conducted three interview sessions with STEM advisors where they participated in group-discussions and think-aloud exercises to reveal how they interacted with Student Explorer. Along with these interview sessions, we conducted weekly meetings with STEM Academy advisors and faculty members. These weekly meetings served as regular points of contact and provided opportunities for advisors to describe how they were using Student Explorer.

During these weekly meetings, advisors provided feedback to us about using Student Explorer as a resource for their individual meetings with students. We asked advisors to describe their process of receiving, opening, and using Student Explorer each week. We were especially interested in the types of information available within Student Explorer that advisors used when deciding to contact STEM Academy students. We also discussed interface design issues, such as what additional data might be useful to advisors and why. For example, we would describe possible sources of data and get their feedback on how they would use these data before and after their advising activities with students. It was important in the design process to know the typical workflow process advisors used so that the most useful data were featured immediately upon opening Student Explorer.

Below, we report our results chronologically, specifying the development of Student Explorer and the ways in which it was used by advisors.

6.4 Development and Use of Student Explorer

6.4.1 Phase I

6.4.1.1 Student Explorer Design

In fall 2010, we began working with the STEM Academy advisors to develop an EWS. During this first phase, we conducted a needs assessment to determine what information would be most useful for advisors to support their advising activities. They reported that the most basic need involved having up-to-date grade information on their students. We provided an initial solution to this problem by querying the campus LMS for all course sites that included a STEM Academy student and that used the GB or Assignments tools (we could not track students' grades unless a course site used either the GB or Assignments tools). We located a large number of courses that fit these criteria, including many of the core engineering and science courses for first- and second-year STEM students. However, some core courses, including first- and second-year mathematics courses, did not use either the GB or Assignments tools. (We were unable to include these courses in Student Explorer due to the lack of LMS data until the most recent iteration of the tool in fall 2012.) In Phase I, we tracked over 150 individual students across 400 courses, with the number of courses per student averaging over 2.6.

Student Percentage Points Earned	Percentage Points behind Course Average	Site Visits Percentile Rank	Classification
>= 85%			Encourage
75% <= X < 85%	< 15%		Explore
75% <= X < 85%	>= 15%	< 25th percentile	Explore
75% <= X < 85%	>= 15%	>= 25th percentile	Encourage
65% <= X < 75%	< 15%	< 25th percentile	Engage
65% <= X < 75%	< 15%	>= 25th percentile	Explore
65% <= X < 75%	>= 15%		Explore
55% <= X < 65%	>= 10%		Explore
55% <= X < 65%	< 10%		Engage
< 55%			Engage

Fig. 6.1 Student explorer classification scheme

The validity of the information provided in the GB and Assignments tools was an additional constraint when working with LMS data. The validity of these data was dependent upon instructors' actual use of the GB and, more generally, use of the LMS as part of their instruction. For example, some instructors used each GB entry to reflect an assignment's contribution to the final course grade (e.g., a ten-point assignment was purposefully meant to account for 10 % of a student's grade out of 100 total points possible for the course). Other instructors applied weights to each GB entry to determine each assignment's contribution to the final course grade. Additionally, instructors varied the actual entries they posted to the GB or Assignments tools. For example, some instructors only posted homework assignments to the GB and did not post grades that contributed substantially to the final course grade, such as exams. Based on instructors' idiosyncratic use of the GB and Assignments tools, we reported a non-grade equivalent percent of points earned out of points possible to academic advisors.

In February 2011, after aggregating GB and Assignments data collected from the LMS, we created a multiple sheet Microsoft Excel file for advisors. We designed an "Advisor Summary" sheet that allowed advisors to view all STEM Academy students' percent of points earned out of points possible for each course in which an STEM Academy student was enrolled. We also created individual sheets for each student–course combination (see Fig. 6.1). These individual sheets provided longitudinal graphical depictions of a student's developing course grade. We provided this file to advisors starting in March 2011 and updated the information on approximately a bi-weekly basis. Near the end of this chapter, we describe some preliminary impacts of Student Explorer by comparing cumulative GPAs of STEM Academy students versus other students in the College of Engineering before and after Student Explorer was used by advisors.

We developed a classification scheme to highlight students who may be in most need of academic support to help advisors parse the large amount of data. By highlighting those students in the greatest need of support, we specified actions that an

advisor could take in relation to that student: Encourage (green), Explore (yellow), and Engage (red). We refined the specific decision rules associated with the classification scheme through three collaboration sessions with all advisors and two interviews with one of the advisors. Classifications were initially generated using two rules: (1) whether a student's percent of points earned was at or above the various non-grade equivalent thresholds (85, 75, or 65 %) and (2) whether a student was a certain percentage below the course average of percent of points earned (10 % at the low end, 15 % at the high end).

Using only these two rules based on GB data, we found that early versions of Student Explorer were oversensitive in classifying students as Explore (yellow) and Engage (red). Student Explorer was particularly oversensitive in the early weeks of a semester when there were few grade entries available. Advisors, however, expressed benefits from over classifying students as Explore (yellow) or Engage (red) based only on a few early course assignments because it provided them with opportunities to hear students describe their own course performances. Classifying these students in such a way provided the opportunity for advisors to identify and provide support to all of these individuals *before* they took their first exam or submitted an assignment that contributed substantially to their final grade. Even though Student Explorer classified more students than were actually performing poorly in some courses, little additional harm came to misclassified students because oversensitivity issues with the classification scheme were attenuated after more points accrued. This was especially true for those courses in which assignments were inherently weighted through the points possible.

While grade information was useful to advisors, we also explored other sources of data from the LMS to help advisors contextualize a student's course grade. We initially thought that providing information about tools that are predictive of final course grades would be beneficial to advisors. For example, we examined correlations between the degree to which students used specific LMS tools, such as Chat and Discussion, and their final course grades. The general strategy of seeking correlations between a tool's use and a final course grade is related to a familiar strategy in LA research—developing prediction models to assess the likelihood of academic failure. In order to examine these patterns, we drew upon LMS data from previous semesters of core courses for College of Engineering students. We used functional data analysis to plot changes in students' use of the LMS in line with their overall course grade. However, we found little evidence that frequently utilized course tools were related to course grades.

Our research efforts surfaced multiple limitations with developing prediction models to assess students' likelihood of academic success. Unlike some systems, such as Intelligent Tutoring Systems that guide and assess student progress through a well-defined problem space, LMSs and a variety of other online learning technologies are necessarily designed to be content agnostic and dependent on how an instructor integrates them into course-specific activities. Given this reality, patterns in LMS use are often not generalizable across multiple course. For example, if the course site is not an important part of course

activities, data from how individual tools are used on these course sites may lead to spurious conclusions related to a student's academic progress. After analyzing patterns across multiple tools and courses, combined with the threats to validity from generalizing across courses, we began relying less on sources of data that are "predictive" and instead incorporated specific sources of data that advisors found useful for understanding a student's course performance and for discussing that performance with them.

6.4.1.2 How Advisors Used Student Explorer

Advisors described their use of Student Explorer during the first phase of the project as "sporadic." Interestingly, early collaboration sessions where we provided advisors mock-ups of Student Explorer designs proved most beneficial to advisors. After looking at the initial displays, drawn from actual student data, advisors were able to identify students who needed immediate help. Advisors contacted these students and worked with them to identify improvement strategies. One reason advisors gave for not using Student Explorer more regularly during this phase is that they had not yet found a way to integrate it into their regular work practices.

Prior to the implementation of Student Explorer, first-year STEM Academy students met with advisors on a monthly basis and turned in progress reports of their course grades. This process continued even after advisors started receiving Student Explorer reports. Advisors have used these progress reports for multiple years and these sheets had become integrated into their regular advising work. Therefore, advisors initially reported that they did not know what value Student Explorer was adding over and above the students' self-generated reports. The extra time it took to view and learn how to use Student Explorer was another reason why advisors did not initially integrate it into their regular work practices. We received a few reports early on from advisors that they did not know how to interpret some of the data or how to make it useful for working with students. After Phase I, we spent more time with the advisors—walking through Student Explorer, describing various features, and discussing possible ways to use the system. As a result, many of these issues did not resurface in Phase II and advisors began to more fully integrate Student Explorer into their advising.

Advisors reported that Student Explorer was most useful in tracking the performance of students in their second or third years who were no longer required to attend monthly advising sessions. While not an initial goal of Student Explorer, tracking these students was useful because they were often "under the radar" until final course grades were made available to advisors. Despite collaboration around the design of the interface, advisors also reported that, overall, Student Explorer was not user friendly, with the exception of the red, yellow, and green color coding. These colors helped advisors quickly make sense of the large amounts of data and identify students in the most need of advising.

Based on the results of this first phase of the project, we found that Student Explorer did not work as we intended but did have some beneficial, unintended effects. Unfortunately, the academic advisors did not find that the system provided added benefits over and above the handwritten progress reports for tracking first-year STEM Academy students. It was, however, useful for identifying low performing students beyond their freshman year—an unintended effect. Specifically, these students were contacted more frequently than in previous semesters as a result of advisor's use of Student Explorer. This increased contact led to more meetings between the advisors and these advanced students, which created more opportunities for these students to avail themselves of academic support services. During the next phase of the design, we worked further with advisors to improve the usability of Student Explorer and to identify ways in which using it led to increased contact with *all* STEM Academy students.

6.4.2 Phase II

6.4.2.1 Improved Student Explorer Design

Throughout the summer of 2011, we worked with advisors to improve the usability of Student Explorer and to identify ways in which using the EWS led to increased contact with all cohorts of the STEM Academy students. We released a new version of Student Explorer to advisors at the start of the fall 2011 semester that included three main changes. First, based on the advisor feedback from the winter 2011 semester, we included labels and grade data for individual assignments on the student report sheets (see Fig. 6.1). While the Engage, Explore, and Encourage classification scheme alerted advisors to those individuals in need of their intervention, it was formerly based solely on relative, intra-course grade measures. As we previously stated, these measures did not distinguish between the importance of GB entries (e.g., tests versus homework assignments), potentially biasing classifications by over-weighting some assignments. The inclusion of individual GB entries specified the source of each grade and helped to clarify for advisors the student's classification within each course. Second, we incorporated the number of times a student accessed a course site into the classification scheme, using it as a proxy for revealing student effort. Specifically, we used a combination of week-to-week and cumulative access events to classify students on the borderline of grade thresholds. For example, if a student was in the bottom quartile for login events (average of week-to-week and cumulative percentile ranks) relative to their peers and was on the borderline between a B and B-, the student would be classified as Explore (yellow) instead of Encourage (green) (see Fig. 6.2). Third, we delivered updated data into the system on a weekly basis and this gave advisors the opportunity to track students and intervene in a timelier manner. We tracked over 200 individual students across 600 different courses in Phase II.

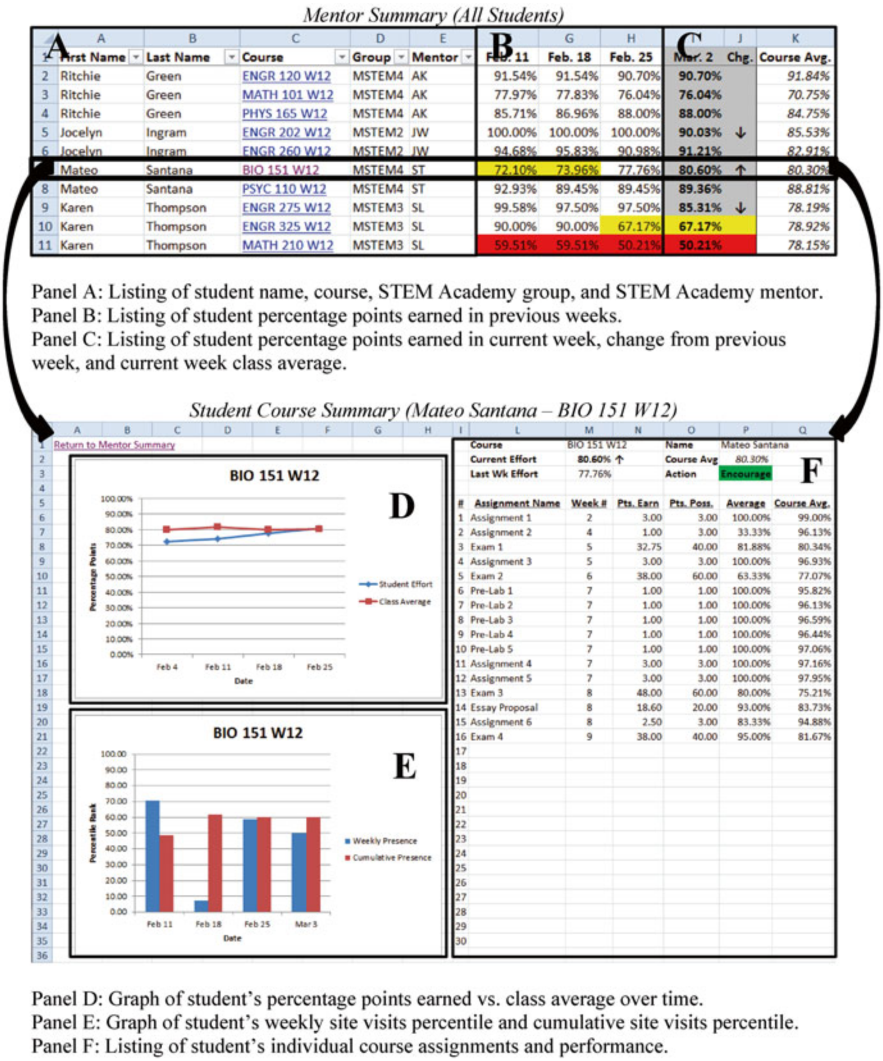

Fig. 6.2 Student explorer screenshots. (**a**) Listing of student name, course, STEM Academy group, and STEM Academy mentor. (**b**) Listing of student percentage points earned in previous weeks. (**c**) Listing of student percentage points earned in current week, change from previous week, and current week class average. (**d**) Graph of student's percentage points earned vs. class average over time. (**e**) Graph of student's weekly site visits percentile and cumulative site visits percentile. (**f**) Listing of student's individual course assignments and performance

6.4.2.2 How Advisors Used Student Explorer

Based on the above design modifications, advisors reported using Student Explorer more frequently than they had in Phase I. Moreover, use of Student Explorer led advisors to request that we incorporate more data from the LMS. For example, after

a few weeks of regular use, one advisor asked us to aggregate and display students' first exam scores for one core engineering course outside of the regular, weekly distributions of the data. Advisors found from previous semesters that the first test in this course was important to a student's final grade. Using this exam data as displayed in Student Explorer, the advisor quickly identified students who did not do well and organized a post-exam session where a graduate student was available to help the students identify what they could do to better perform on the next test.

Advisors found the new design feature of showing the labels and grade data for individual GB entries particularly useful for focusing their interactions with students. Having the labels and grades for individual GB entries allowed advisors to address specific areas of concern within a course as it related to a student's overall performance in the course. For example, one advisor specifically targeted students' performances on major exams to help students make decisions about dropping courses, and to discuss the degree to which a student needed to work more closely with the course instructor or teaching assistant to improve his or her grade.

We were also able to identify tentative links between the ways in which advisors used Student Explorer and the frequency with which advisors contacted students. In Phase II, advisors reported contacting all students classified as Engage (red) on a weekly basis. Advisors typically emailed students to schedule a meeting to discuss specific improvement strategies. Advisors reported receiving immediate responses from all first-year STEM Academy students and from approximately half of all STEM Academy students who had been in the program longer than 1 year. Suggested strategies for student improvement included setting up appointments with tutoring services, meeting with a course instructor, attending office hours, and reading through the instructor's posted PowerPoint slides before and after a lecture. Thus, Student Explorer provided a view of student performance that signaled to advisors which students they might want to contact and what types of issues needed to be addressed. Use of Student Explorer also resulted in a more immediate opportunity to suggest various improvement strategies rather than waiting for regularly scheduled meetings and relying on students' self-reports.

In what follows, we describe how one advisor used the Student Explorer (1) filter the spreadsheet to identify assigned students, (2) locate all students who were classified as Engage (red), and (3) view each student's individual grade sheet. After selecting an individual student's sheet, (4) the advisor would then examine how the student performed on each grade entry. After examining individual entries, (5) the advisor would examine a longitudinal graph of the student's course grade. The advisor would then (6) email the student requesting a meeting to discuss his or her academic performance. Though individual advisors varied in the processes by which they used Student Explorer and contacted students, in general, advisors stated that they used multiple sources of additional information in conjunction with the data presented in Student Explorer. For example, prior personal and academic histories for each student and the advisor's own knowledge about specific courses and instructors impacted what they would say in both their initial communications and one-on-one meetings with a student.

6.4.2.3 Potential Outcomes of Student Explorer Use

In Phase II, we assessed potential outcomes associated with advisors' use of Student Explorer. For one outcome, we examined the number of first-year students who were likely contacted between their regularly scheduled monthly meetings. The purpose of this analysis was to see whether Student Explorer increased the frequency of advisors' contact with students. Based on the feedback that advisors gave during interviews, students were contacted by their advisor through e-mail when they were classified as Engage (red). Based on this feedback, we examined three different weeks throughout the semester where no student had a regularly scheduled meeting. These weeks represented opportunities for advisors to contact students that were not present before implementation of Student Explorer. In the first intervening week, 2 of the first-year STEM Academy students and 25 of the second through fourth year students were classified as Engage. In the second intervening week, 3 of the first year students and 23 of the second-fourth year students were classified as Engage, respectively. In the third intervening week, 7 of the first year students and 27 of the second-fourth year students were classified as Engage. Because advisors reported additional contact with all students classified as Engage, these numbers reflect a potential overall increase in the number of times students were contacted by an advisor between regularly scheduled meetings.

Along with identifying whether or not advisors contacted students more frequently as a result of using Student Explorer, we compared cohorts of students before and after advisors implemented the system. For comparison purposes, we selected students' sophomore grade point averages (GPAs); STEM Academy students' sophomore year represented the first year that they did not regularly meet with their advisors, which created an opportunity to see the effects of increased communication between advisors and students due to Student Explorer use and not monthly meetings. We speculate that the effects of increased communication would be most noticeable in this key transition year. For this analysis, we compared the means and interquartile ranges for STEM Academy sophomores prior to the adoption of Student Explorer (2008–2009 and 2009–2010 academic years, shown on Fig. 6.3 in red) against STEM Academy sophomores following the adoption of Student Explorer (2010–2011 and 2011–2012 academic years, in Fig. 6.3 in blue). We further drilled down into the data by examining differences based on students' incoming math ACT scores; we created two groups of students—those who scored *below* the median of incoming STEM Academy students' math ACT scores (Fig. 6.3, left) and those who scored *above* the median (Fig. 6.3, right). As illustrated in Fig. 6.3, across both ACT groups, there were overall increases in sophomore GPAs after Student Explorer adoption. While these data are suggestive, they may also reflect general changes in who got admitted to STEM Academy; we observed a 2-point increase in average ACT scores from 2008 to 2012. We do note, however, that after advisors' use of Student Explorer, most sophomores achieved a GPA at or above 3.0, which was a primary goal for students in the Academy.

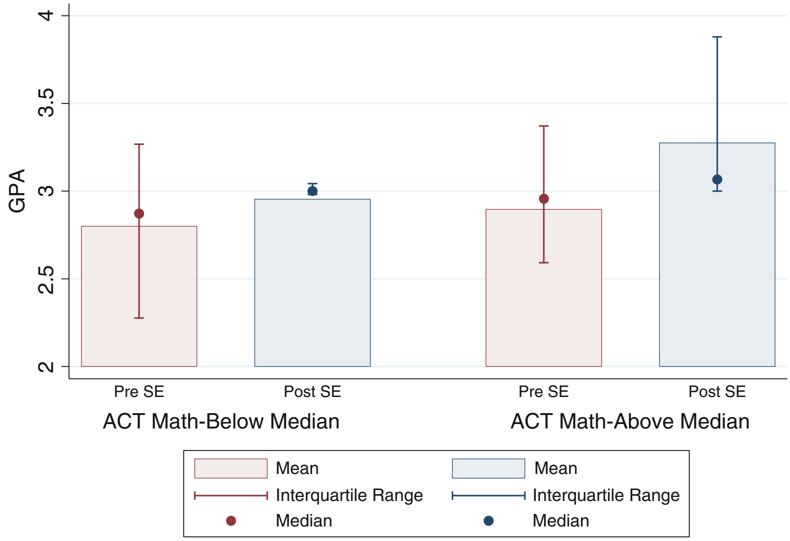

Fig. 6.3 Differences between STEM Academy sophomore GPAs before and after student explorer implementation

The above analyses are examples of formative assessments used by the research team and STEM Academy collaborators in Phase II. The purpose of these analyses were to create opportunities to have further discussions among the research team and to seed design decisions and speculate on possible supports needed by STEM Academy advisors. One of our goals for future work with Student Explorer is to develop mechanisms for tracking when and how students follow-up on suggestions provided by their academic advisors. We view the GPA data in formative terms, providing potential rationales for ongoing design decisions to be explored collaboratively with STEM Academy advisors.

6.5 Conclusion

This chapter reported on two phases of a multi-year project aimed at developing an EWS, Student Explorer, for an undergraduate engineering advising program. The advising program served as a strategic research partner providing an authentic context in which to explore important issues related to LA-based interventions. Moreover, the STEM Academy simultaneously provided the opportunity to develop a working product that supported Academy advisors identify students in need of academic support.

Interactions among intermediaries (e.g., academic advisors), students, and an EWS are complex and extend over time—breakdowns within and between interactions can lead to less than desired outcomes. In this project, we developed Student

Explorer to influence how an advisor identifies students in need of academic support, with the intent of increasing the frequency with which an advisor contacts students and engages with them in discussions around their academic performances. An important step for LA research more generally will be to further clarify the complex relationships between providing data to interested parties, such as academic advisors and instructors, and improved student outcomes. Generally speaking, there are multiple, possible breakdowns between providing actionable data and positive outcomes. For example, in our research, we described how STEM Academy advisors received and made sense of data on hundreds of students across hundreds of courses. Advisors contacted students and offered recommendations, which students may or may not have acted on. Depending upon what occurs following a recommendation, a student might have performed better than expected on a subsequent course assessment.

Since Phase II, the design of Student Explorer has refined, expanded, made available online, and much of the data processing has been automated. For a discussion of lessons learned about scaling from research project-level implementation to university-level IT support, see Lonn et al. (2013). As Student Explorer was scaled a variety of challenges have emerged, including differing approaches to advising and perspectives on the utility of EWS data for understanding how, when, and why students' academic performance may be declining. As we address these challenges, we are cognizant of the need to continually communicate with advisors about how they interpret and make recommendations based on the data presented in Student Explorer. Future LA research will benefit from identifying the ways in which system developers, researchers, and data scientists support key actors (advisors, instructors, students) in making sense of and acting on data generated by LA-based interventions.

References

Barab, S. A., & Squire, K. D. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*(1), 1–14.

Beck, H. P., & Davidson, W. D. (2001). Establishing an early warning system: Predicting low grades in college students from survey of academic orientations scores. *Research in Higher Education, 42*(6), 709–723.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178.

Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *Educause Review, 42*(4), 40–57. Retrieved from http://net.educause.edu/ir/library/pdf/ERM0742.pdf

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13. 35–37.

Collins, A. M. (1992). Towards a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin: Springer.

Collins, A. M., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15–42.

Dahlstrom, E., de Boor, T., Grunwald, P., & Vockley, M. (2011). *The ECAR national study of undergraduate students and information technology, 2011*. Boulder, CO: EDUCAUSE.

Dawson, S., McWilliam, E., & Tan, J. P. L. (2008). Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. *Hello! Where Are You in the Landscape of Educational technology? Proceedings Ascilite* (pp. 221–230). Melbourne. Retrieved from http://www.ascilite.org.au/conferences/melbourne08/procs/dawson.pdf

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. *Proceedings of the First International Conference on Learning Analytics and Knowledge* (pp. 9–17). Banff, Alberta, Canada: ACM.

Fritz, J. (2011). Classroom walls that talk: Using online course activity data of successful students to raise self- awareness of underperforming peers. *The Internet and Higher Education, 14*(2), 89–97.

Goggins, S., Galyen, K., & Laffey, J. (2010). Network analysis of trace data for the support of group work: Activity patterns in a completely online course. *Proceedings of the 16th ACM International Conference on Supporting Group Work* (pp. 107–116). Sanibel Island, FL.

Goldstein, P., & Katz, R. (2005). *Academic analytics: The uses of management information and technology in higher education—Key findings (key findings)* (pp. 1–12). EDUCAUSE Center for Applied Research. http://www.educause.edu/ECAR/AcademicAnalyticsTheUsesofMana/156526

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making (NCEE 2009-4067)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K., (2011). *The 2011 Horizon report.* Austin, TX: The New Media Consortium. Retrieved from http://wp.nmc.org/horizon2011/

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). New York: Cambridge University Press.

Lonn, S., Aguilar, S., & Teasley, S. D. (2013). Issues, challenges, and lessons learned when scaling up a learning analytics intervention. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 235–239). Leuven, Belgium: ACM.

Lonn, S., Krumm, A. E., Waddington, R. J., & Teasley, S. D. (2012). Bridging the gap from knowledge to action: Putting analytics in the hands of academic advisors. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 184–187). Vancouver, Canada: ACM.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599.

Maton, K. I., Hrabowski, F. A., III, & Schmitt, C. L. (2000). African American college students excelling in the sciences: College and postcollege outcomes in the Meyerhoff Scholars Program. *Journal of Research in Science Teaching, 37*(7), 629–654.

Matsui, J., Liu, R., & Kane, C. M. (2003). Evaluating a science diversity program at UC Berkeley: More questions than answers. *Cell Biology Education, 2*(2), 113–121. doi:10.1187/cbe.02-10-0050.

May, M., George, S., & Prévôt, P. (2011). TrAVis to enhance online tutoring and learning activities: Real-time visualization of students tracking data. *Interactive Technology and Smart Education, 8*(1), 52–69.

McKay, T., Miller, K., & Tritz, J. (2012). *What to do with actionable intelligence: E2Coach as an intervention engine*. Paper presented at The 2nd International Conference on Learning Analytics and Knowledge. Vancouver, Canada.

Morris, L. V., Finnegan, C., & Wu, S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education, 8*(3), 221–231.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and Matlab. Use R series*. New York: Springer.

Part III
Learning Analytics for Teachers and Learners

Chapter 7
The Data-Assisted Approach to Building Intelligent Technology-Enhanced Learning Environments

Christopher Brooks, Jim Greer, and Carl Gutwin

7.1 Introduction

The purpose of this chapter is to describe the data-assisted approach to embedding intelligence in technology-enhanced learning environments that leverages the sensemaking process of instructional experts. In this approach, instructional experts are provided with summaries of the activities of learners who interact with technology-enhanced learning tools. These experts, which may include instructors, instructional designers, educational technologists, and others, use this data to gain insight into the activities of their learners. These insights lead experts to form instructional interventions that can be used to enhance the learning experience. The novel aspect of this approach is that it frames the learning environment as a system that is not just made up of learners and software constructs, but also of the educational experts who may be supporting the learning process. This approach demonstrates how the sensemaking process in the field of learning analytics can be used to affect teaching and learning.

Higher education increasingly makes use of courses with large cohorts of learners and smaller instructor-to-learner ratios. Bloom (1984) demonstrated that learners who are taught in one-on-one learning have, on average, summative assessment marks two standard deviations higher than those taught in a traditional classroom setting. This finding is a principal motivator in the academic research area of *intelligent tutoring systems* (ITS), where software systems form models of learners and adapt the learning environment based on learner performance, much the way a human tutor would. Despite several commercial successes (e.g. Carnegie Learning 1998; Suraweera and Mitrovic 2002) and continued

C. Brooks (✉)
School of Information, University of Michigan, Ann Arbor, MI, USA
e-mail: brooksch@umich.edu

J. Greer • C. Gutwin
Department of Computer Science, University of Saskatchewan, SK, Canada

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_7, © Springer Science+Business Media New York 2014

development of the research field[1], ITS have been deployed only in a few isolated cases and have introduced minimal changes in day-to-day teaching and learning in higher education. In part this is an effect of the expense and labour required to build an intelligent tutor: it requires the work of a domain expert to outline how concepts relate to one another, a pedagogical expert to provide a method for determining when a learner has made a mistake and what should be done to mediate the issue, and a content expert to build initial and remedial content to be delivered to learners. Instructors in higher education rarely have such resources available to them, and instead have to rely upon their own understanding of each of the aforementioned areas, understanding which may be limited and which may be difficult to implement as an automated computational tutor.

Instead of ITS, universities and colleges have focused on scalability with learning technologies and have widely deployed *learning management systems*. These solutions include content creation and delivery solutions, synchronous and asynchronous discussion forums, and multimedia streaming, capture, and playback systems, and provide infrastructure support for learning activities. Learning management systems are used to augment traditional teaching and learning experiences as well as to provide distance and online learning. The environments are, in a sense, non-ITS, where personalization of course content is difficult and the instructor is relied upon to make an a priori identification of the kinds of problems learners may have. While these benefit from a reduction in the number of experienced programmers and knowledge engineers needed, helping scalability and thus leading to greater adoption in higher education, these systems suffer from some of the same resource issues that affect ITS. In order to ensure that all learners are supported, a breadth of content needs to be developed that fits the expected needs and goals of those learners. In sufficiently large or diverse courses, this results in a team of instructional design experts being used to build a comprehensive and pedagogically sound course offering. The end result is the same in terms of a priori effort: building instructional interventions requires significant up-front modelling of learners, pedagogical approaches, and domain-specific content.

In most technology-enhanced learning situations, the interactions happen between the learner and the learning environment, and instructional experts are often limited in their ability to see these interactions and make necessary interventions. Instead of the traditional method of relying heavily on a team of experts who have mapped out the space of possible challenges, difficulties, and misconceptions learners might encounter a priori, we argue that the instructional experts involved in delivering a course can use learner interaction data and make use of their contextual knowledge of the content, cohort, and pedagogy to provide a more individualized learning experience. By considering the issues that learners face while the course is being offered, a smaller highly contextual problem space (e.g. how to teach a specific concept that a specific group of learners are having issues with) need be considered instead of a broader

---

[1] In particular see the *Intelligent Tutoring Systems* (*ITS*) and *Artificial Intelligence in Education* (*AIED*) conference series, as well as the *Journal of User Modeling and User-Adapted Interaction* (*UMUAI*).

general problem space (e.g. the design of a whole course for learners with a variety of backgrounds). Instructors often employ pedagogical interventions as a part of their normal teaching practice in traditional lectures—it is well known that master teachers observe the behaviours of their students and adapt instruction on demand (Buskist 2004)—but as course sizes grow and interaction moves online, both the observation of social cues that inform the instructor and the ability of that instructor to intervene based on those behaviours becomes more difficult. In this chapter we will show how making the hidden behaviours of learners visible to the instructional expert allows him or her to form *insights*, and enables him or her to react to these insights from technology-enhanced learning environments with *instructional interventions*.

The key aspect of the data-assisted approach to supporting instructional interventions in technology-enhanced learning environments is that both the insights and the interventions are the result of a dialogue between the intelligence in the system and intelligence of the instructional expert. Traditionally there are two forms of intelligent learning environments: those that are *adaptive*, and those that are *adaptable*. Adaptive environments are those described as systems that automatically change in response to user activity based on some pre-programmed knowledge base and rule set. This classification places intelligence in the software system itself. Adaptable environments are those that are personalizable based on direct requests from the learner. This represents a view that the intelligence exists in the learner alone. The continuum between these two approaches is referred to as the *locus of control*, and historically has placed artificial intelligence techniques at one end and human–computer interaction techniques at the other. In the data-assisted approach, interaction information is collected about learners and made available to a third kind of actor, the instructional expert. This expert goes through a sensemaking process, which may be while a course is being run or in between courses, and acts on insights they find by providing pedagogical interventions in the learning system.

The next section of this chapter will describe the data-assisted approach in more detail, ending with three motivating scenarios describing how it can be applied. Each of these scenarios purposefully considers different instructional expert roles (instructors, instructional designers, and educational technologists) and the reasons they might interact with the technology-enhanced learning system. Aspects of these scenarios have each been investigated in depth, and Sect. 7.3 provides case studies demonstrating the outcomes of applying the data-assisted approach. This is followed with a brief conclusion of our findings in realizing this approach.

7.2 The Data-Assisted Approach

The data-assisted approach is intended to generate *insight* that leads to and supports *instructional interventions*. It is different from traditional methods of building intelligent educational systems in that it explicitly acknowledges the role of instructional experts. It is through these experts that insight is generated, and from these experts that instructional interventions come.

Before talking about methods of generating insight or interventions, it is useful to talk about the experts themselves. Instructional experts for a particular environment can be many different people with different tasks. These experts can be instructors, instructional designers, content creators, tutors, markers, or peer learners. The experts can have a formal role in the course, or can be informal actors that engage with the environment through happen-stance. There is already a body of work that examines how learners gain insights about their own actions (e.g. open or scrutable learner modelling), as well as how learners can adapt an environment to their own needs (e.g. adaptable systems). But how instructors tune learning environments through a sensemaking process has been largely unaddressed by the technology-enhanced learning environment community.

The data-assisted approach starts with the collection of information about learners as they interact with tools in the learning environment. This information describes the behaviours of learners, and may be collected through a variety of means across a variety of different tools (see Sect. 7.2.1). The data must then be summarized and correlated to identify groups of learners based on some educational attributes, tasks, or goals.

The process of aggregating and correlating data with pedagogical goals requires that the instructional experts interact with computational elements of the system. A number of different computational approaches can be used, and this chapter will demonstrate how two approaches in particular, *information visualization* and *unsupervised machine learning*, can be successful in identifying aggregations of learners. Regardless of the computational technique involved, it is the instructional expert who gives meaning to the groups, and identifies or parameterizes appropriate instructional interventions.

One of the differences between the data-assisted approach and traditional intelligent learning environments is that the interventions are largely based on the expertise of the instructional expert and not on a priori domain, curriculum, or pedagogical knowledge that has been formalized and loaded into the system. The intelligence is in the overall system that includes the learners, the software, and the instructional expert, and not in software alone. This does not forbid traditional intelligent software (e.g. ITS) from interacting inside a data-assisted approach, of course, but changes the focus of intelligence to be system-wide instead of being located in software alone.

As the data-assisted approach relies upon the instructional expert to provide pedagogical knowledge, it is compatible with many learning theories and instructional design approaches. For instance, an instructor who observes a deficiency of learning in one set of learners might decide that a social constructivist approach is appropriate as a particular intervention, and perhaps configure an online learning environment to send learners to a blogging activity. Another instructor might, after seeing the same grouping of learners, modify the curriculum to include another assignment that focuses on experiential learning. Regardless, both instructors can use the data-assisted approach to get insights into student activity and to customize the course for those learners.

7.2.1 *What Is Meant by* Data *in the Data-Assisted Approach?*

As learners interact with educational environments, traces of their activities can be logged. These traces link actors (e.g. learners, instructors, instructional assistants) and artefacts (e.g. videos, tests, web pages) with interaction behaviours (e.g. watching, answering, clicking). Most educational environments already have some form of trace logging, although many don't separate functional traces (those that pertain to the correct operation of the system) from informational traces (those that are useful specifically for data analysis).

The granularity of the data collected is an important factor to consider when creating an educational environment. One approach is to capture low-level interactions such as key presses or mouse movements (sometimes referred to as clickstreams (McCalla 2004; Peckham and McCalla 2012)), while another is to use more coarse-grained events such as whether or not an answer to a question was correct. The choice between a fine or coarse level of granularity affects the types of behaviours that are available for analysis. A fine level of granularity (e.g. individual key presses) may require the end-user to aggregate data in order to form meaningful insights that they believe are useful (e.g. concepts in a course), while a coarse level of granularity (e.g. a grade on an online examination) may not be decomposable in order to relate to the same meaningful insight. A number of pragmatic concerns around the capture and storage of user data also exists—collecting mouse pointer data as a user interacts with an online content management system, for instance, would create a significant amount of data that would need to be transferred back to a central location for storage. In addition, this level of trace data would almost certainly need to be summarized in order to be useful as an attribute in either automated reasoning or visualization techniques.

A number of authors have considered how semantics can be added to learner trace data to make analysis easier, including (Najjar and Wolpers 2006; Brooks et al. 2004). Building tools to support the data-assisted approach requires a consideration of how data is collected and labelled such that it is meaningful, but this activity takes place outside of the adaptation process, and is largely one of traditional knowledge engineering and system design.

7.2.2 *What Does It Mean to* Assist *in the Data-Assisted Approach?*

The summarization of learner traces into meaningful insights can be considered a form of learner modelling, where the underlying data for the modelling process comes from the interactions learners have had with the learning environment, and the model itself is represented by the insights that are generated from this data. An important differentiator between the data-assisted approach and other forms of

learner modelling is that in the data-assisted approach the instructional expert is considered the key actor in the modelling processes. Instead of loading instructional intelligence into the software a priori, there is a reliance on the instructional expert to form hypotheses of learning activity, validate the pedagogical relevance of patterns, and form instructional interventions as appropriate. Thus, software to support data-assisted investigation must collect data about learners, allow instructors to parameterize the analysis of this data as needed, summarize and present patterns of behaviours to the instructors, and map instructional interventions to groups of learners as instructed. The intelligence in such software may be quite limited depending on the data being collected and the mechanisms being used to present this to instructors; on one end of the spectrum, a data-assisted system may rely totally on information visualization techniques with minimal filtering to provide a summary to the instructor, while at the other end of the spectrum a data-assisted approach may provide sophisticated artificial intelligence techniques such as unsupervised machine learning mechanisms to aggregate learners into cliques. It is through the application of these techniques that assistance is given to instructional experts to complete the modelling process.

Embedding the instructor in the learner modelling process aims to increase the scalability of adaptive e-learning systems. While current intelligent learning environments scale well to many learners, they do not scale well between domains, and require a significant amount of domain and pedagogy modelling when being applied to new curricula. By helping instructors to form groups of learners based on behaviours, the data-assisted approach seeks to be a generalizable approach to building adaptive learning environments. Human intelligence in the form of the instructor is used to provide the labelling of groups and their relationships to adaptive components, while artificial intelligence and information visualization can be used to provide a statistical and graphical understanding of the relationships between observed behaviours and groups of learners.

The use of human intelligence in the data-assisted approach fits well with the notion of *sensemaking* as a principal goal of the field of learning analytics as described by Siemens (2012). Klein et al. (2006) describe sensemaking as a process by which events can be understood with consideration of perspectives, which they refer to as a *frames*:

> We can express frames in various meaningful forms, including stories, maps, organizational diagrams, or scripts, and can use them in subsequent and parallel processes. Even though frames define what count as data, they themselves actually shape the data (for example, a house fire will be perceived differently by the homeowner, the firefighters, and the arson investigators). Furthermore, frames change as we acquire data. In other words, this is a two way street: Frames shape and define the relevant data, and data mandate that frames change in nontrivial ways. (Page 88 of Klein et al. 2006)

The process of the instructional expert interacting with the data collection and analysis aspects of the learning environment can be thought of as a dialogue; as the instructional expert elicits the formation of groups or classifications of learners from the system, he or she can form new hypotheses as to the state of learning happening in each group, and modify how course attributes such as content, sequencing,

activities, or tools are made available. The dialogue is bidirectional—the instructor uses the environment to better understand learners, and shares this understanding of learners with the system through labelling. The system takes these labels, and the associated adaptations provided, and applies them to learners who fall within the groups based on learner behaviours.

The dialogue between system and instructional expert is also an ongoing one. As learners continue to interact, or as new learners begin to interact, with the tools in a learning environment, new data and new behaviours can be presented to the instructional expert. This expert creates, deletes, and modifies instructional interventions as they apply to learning objectives and pedagogical approach. Thus, the modelling process is intended to fit within the day-to-day activities of an instructional expert such as a course instructor, and not be an activity that takes place before learners are present, as much traditional instructional design activity does. This is not to say there is no place for a priori design in electronic learning environments or that, when available, resources to build ITS should not be utilized. Instead, the suggestion here is that the data-assisted approach is a complementary method to both instructional design activities and intelligent tutoring methods, and requires a different kind of resource that may be more readily available in the context of higher education (e.g. the instructional expert).

7.2.3 Motivating Scenarios

The rest of this chapter will consider how the data-assisted approach might be applied in three different real-world contexts. To aid in this, we present here three scenarios that explore the needs of different kinds of instructional experts. While the scenarios presented here are fabricated, each will be paired with details from actual investigations (Sect. 7.3). The emphasis of these case studies is not on completely satisfying the requirements of each scenario (a monumental task), but in providing evidence that demonstrates how the data-assisted approach has been used to solve particular problems.

7.2.3.1 Scenario One: Visualizing Community Interactions

Katheryn is a faculty member who regularly teaches introductory Computer Science for non-majors. This course is typically moderate in size (50–100 students) and made up of learners from a wide variety of disciplines. This year, Katheryn is teaching the course in an online capacity instead of in a traditional lecture format. Learners have access to an online content management system that includes sets of web pages describing content as well as an asynchronous discussion forum. There are 20 learners enrolled, and because of the distance component the majority of evaluation is weighted on the final examination and assignments.

The content for the course is broad in nature, and Katheryn feels that encouraging group discussion will be one key to keeping learners engaged. She is concerned that the distance modality of the course will cause learners to shy away from interpersonal interactions, and scaffolds this aspect of the course by making weekly reading assignments. Each week, she will post questions to the discussion forums and learners will have to reply with their thoughts on the issues. By having learners read one another's posts, she hopes they will form a shared sense of community, resulting in greater engagement and deeper learning.

In a traditional discussion forum system, Katheryn can see only that students have written messages, as well as the content of those messages. All other interactions, such as traces related to the reading of postings, are inaccessible to her. Using the data-assisted approach, tools can be built to visualize these hidden interactions, enabling Katheryn to more deeply understand how learners are collaborating.

An important aspect of the data-assisted approach is that the underlying data is summarized as it is presented to the instructional expert, and the expert is actively involved in the summarization process. The intelligence in the system is thus a *mediated* and *emergent* property; it exists because of the interaction between the domain and pedagogy expert and the data-rich learning environment. In this scenario, Katheryn has some specific thoughts as to what active engagement means; i.e. that learners are reading and considering one another's messages. Katheryn might be able to aid in the summary of the data by modifying the visualization to render traces differently based on her goals. She also might be able to compare her current class visualizations with those taken in previous years, or those of other classes where she knows different pedagogical techniques are being employed.

The data-assisted approach for supporting instructional interventions is made up of two activities: the summarization of usage data to generate insight, and the support for making instructional interventions based on this insight. This scenario has focused on the first of these, and demonstrates that by revealing to instructors the hidden behaviours of learners (such as the reading of a message), an instructor can understand the affects of their pedagogical practice. The data-assisted approach does not require that instructional interventions take place directly in the technology-enhanced learning environment; insights can be leveraged to form traditional instructional interventions, such as curriculum changes or changes to assignment activities (as Katheryn did). By providing mechanisms to compare visualizations over time, instructional experts can compare the effects of their actions and deepen their understanding of the learning cohort as well as their pedagogical practice.

7.2.3.2 Scenario Two: Measuring Educational Impact

Michelle is an instructional designer supporting a second year undergraduate Chemistry course. This course serves as a service course for other colleges (Agriculture, Engineering, Medicine) as well as for the core Chemistry programme. The course is taught by five different instructors with a common curriculum and common set of examinations. The course reaches a total of 600 learners in a given semester. One instructor has agreed to have his lectures recorded using

a lecture capture and playback system. These recordings are then made available (asynchronously) to all students enrolled in the course.

Michelle's responsibility with the course is to manage both the short-term and long-term instructional design. The lecture recording system is new this year, and Michelle is interested in better understanding how learners use recorded lectures and the effect they have on performance. In particular, Michelle is interested in understanding whether the students who use the lecture capture system incorporate it regularly into their study habits, or whether they just use it only during the examination period. If there is a difference in performance between these groups she will adjust the course content as appropriate, guiding the learners to more effective study habits.

The data-assisted approach is ideal for Michelle as she is interested in exploring the effect of introducing a new tool as well as tuning the course to take best advantage of that tool. Michelle has deep knowledge of different instructional interventions, but lacks an awareness of the underlying activities of learners on which to base these. In a traditional classroom Michelle could go to lectures and observe attendance, a time-consuming process. Further, the lack of time shifting tools in a traditional setting (such as lecture recording) can cause learners to rely on traditional methods for consuming content (e.g. notes or textbooks) that Michelle cannot easily observe. With the data-assisted approach, learning tools are augmented with tracking features that log the activities of learners and make them available to experts like Michelle. While there may still exist modalities of learning outside these, the set of tools that collect learner traces along with methods of summarizing and acting upon these traces form the basis of the learning environment. Michelle is now able to watch in real-time as learners use the system, and leverage her pedagogical expertise to change the way in which the course is designed. Further, Michelle is able to use these traces of learner activity to expand her understanding of the knowledge domain and how learners interact with it.

Michelle's interest in seeing groups of related students fits well with unsupervised machine learning techniques (*clustering*). For instance, Michelle might be interested in grouping learners by their weekly access patterns in the lecture capture system and comparing these to midterm evaluation marks. Depending on her ability, she might directly interact with the logging subsystem of the lecture capture product, or a user-friendly graphical interface (e.g. as in Brooks et al. 2012) may provide her the ability to choose the kinds of data she is interested in clustering. Regardless of the underlying tool used, as the system returns results for her to consider, she may tweak her clustering parameters and attributes based on her background knowledge of what educational theory is appropriate.

This scenario has illuminated two aspects of the teaching and learning environment that are relevant to the data-assisted approach. First, courses are often large multi-section offerings made up of hundreds of learners where a visual in-person analysis of activity is either difficult or impossible. As learning opportunities increasingly happen outside of the classroom (e.g. through time shifting of lectures using lecture capture), the need to capture interaction between learners and technology grows. The data-assisted approach is built on the premise that these interactions allow instructional experts to gain valuable insights into how learners interact with the educational environment.

Second, the data-assisted approach puts the instructional expert at the centre of the *sensemaking* process. In this scenario, it is Michelle who determines whether the clusters fit with her pedagogical approach and are relevant to her investigation. The educational environment helps aggregate and correlate behaviours, but it is the expert who validates the results and forms interventions. How these findings might be tied to interventions is largely omitted from this scenario, but will be touched upon in the next scenario (Sect. 7.2.3.3).

7.2.3.3 Scenario Three: Adapting Learning Environments to Tasks

The previous scenarios have described how an instructor and an instructional designer might use data-assisted approach tools in their daily activities. Intelligent learning environments typically offer automatic adaptivity in response to learner behaviours. The data-assisted approach is appropriate for configuring and building of these systems by bringing technologists into the sensemaking process. Whereas in a traditional intelligent learning environment the sensemaking is done a priori by technologists, a data-assisted intelligent learning environment encourages the creation of adaptive features, or the configuration of existing adaptive features, in response to new patterns of behaviours observed.

This scenario follows Adam, an educational technologist with a software engineering background, who is supporting a lecture capture system similar in nature to the one Michelle is using. Adam has received a lot of feedback from instructors that the navigation in the application is hard for students, especially since the principal method of navigation through the video is by selecting from a series of index thumbnails that are generated from the captured data projector every 5 min of recording (Fig. 7.1). Some instructors want index thumbnails more often, while others want fewer thumbnails that are more like chapter markers in traditional DVD media.

Adam understands the perspectives the instructors have shared, and he believes that students navigate through the videos according to their needs. He may examine the behaviours of learners using visualization tools for the lecture capture system that have similar capabilities as the ones Katheryn used in Sect. 7.2.3.1. Students rarely use the same indices when navigating, but he could use clustering tools (such as described in Sect. 7.2.3.2) to find three clusters of activity which he might label as *reviewing* for frequent use of many indices, *watching* for those students who seem to only use a few indices that refer to significant breaks in the lecture, and *images* for a cluster of access that appears to focus on images in slide content (e.g. chemical drawings in material sciences courses, paintings shown in art history courses, or diagrams used in paediatric nursing courses). He may then either parameterize the indexing functionality of the lecture capture environment based on these clusters, or change the functionality of the system as appropriate.

In this scenario, Adam has used the data-assisted approach to understand how learners are using the lecture recording tools. He has identified patterns of behaviour using tools similar to those used by Katheryn and Michelle. As shown in this section, the data-assisted approach does not have to end with just the detection of patterns.

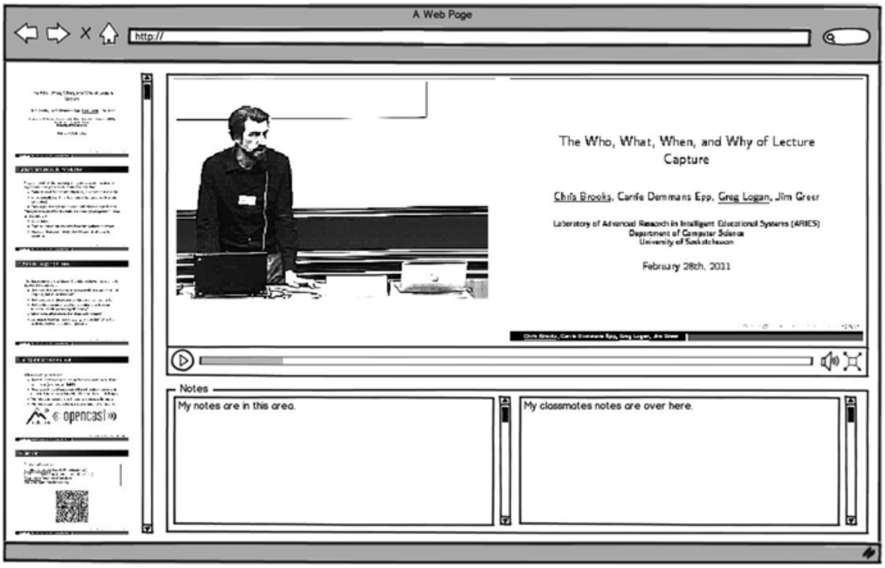

Fig. 7.1 A mock up of a lecture capture environment. The images down the *left hand side* control navigation through the video, and allow a user to seek to a particular *chapter* directly. The video playback component on the *upper right* includes video of both the instructor and the data projector, as well as a traditional scrubber that allows for non-chapter seeking. Additional tools such as note-taking components, discussion forums, or suggested readings are available underneath the video

While Katheryn and Michelle were able to leverage insights to change their pedagogical approach and instructional designs respectively, Adam is making the results of the patterns available to the educational environment by labelling clusters. With labels attached to prototypical behaviours (the student *centroids*), learners can be classified into groups and the appropriate indices can be shown automatically.

7.2.4 Conclusion

Here we have described how the data-assisted approach works, and how it might be used to aid in teaching and learning process. The scenarios presented here are intended to be motivational, and describe how different instructional experts might interact with learning environments that collect end-user behaviour data. To this end, a number of different actors (instructors, instructional designers, and educational technologists) with different use cases (supporting learners, analyzing effectiveness of tools, and customizing the learning environment) using different instructional modalities (online distance learning and traditional face-to-face lecture learning) teaching to different-sized class have been considered.

7.3 Case Studies

7.3.1 Introduction

Over the last 7 years, the laboratory for Advanced Research in Intelligent Learning Environments (ARIES) at the University of Saskatchewan has been applying the data-assisted approach to different technology-enhanced learning environments. The goal here has been to validate that the approach is appropriate for enabling new technology, as well as to realize the teaching and learning gains of these next-generation environments. To this end, thousands of higher education learners from a variety of disciplines have taken part in studies that have used the tools we will describe.

Due to the scope and depth of the studies we have engaged in, a full account of the findings from these studies would be difficult to provide here. Instead, we focus this section on describing three cases that illuminate how the data-assisted approach has engaged experts similar to those presented in the previous scenarios.

7.3.2 Case Study One: Visualizing Community Interactions

One way the data-assisted approach is different from other methods of creating intelligent learning environments is in the way insight is generated. In the data-assisted approach, insight is created through an interaction between an instructional expert and data that represents a particular cohort of interest (a group of learners). A more traditional intelligent learning environment approach would be to fully describe the learning space before the system is deployed, then categorize learners and react accordingly. The data-assisted approach is more reactive, and focuses on in situ exploration of learner activities.

Similarly, instructional interventions can be constructed and delivered through multiple methods. A traditional intelligent learning environment such as an ITS or adaptive hypermedia system changes content delivered to a learner in reaction to their activity. These systems will not often react by changing pedagogical approach, however, unless they have been specifically pre-programmed to do so. In contrast, human instructors often react by considering how they might restructure aspects of a course. By giving these experts control over how instructional interventions are formed, larger pedagogical changes can be made with minimal a priori consideration.

Based on these differences, this section explores two questions that arise when considering the scenario involving Katheryn in Sect. 7.2.3.1:

1. Is it possible to augment a discussion forum system to capture the hidden traces?
2. Can an instructor derive enough meaning from a visualization that they are able to modify or improve upon their pedagogical practice?

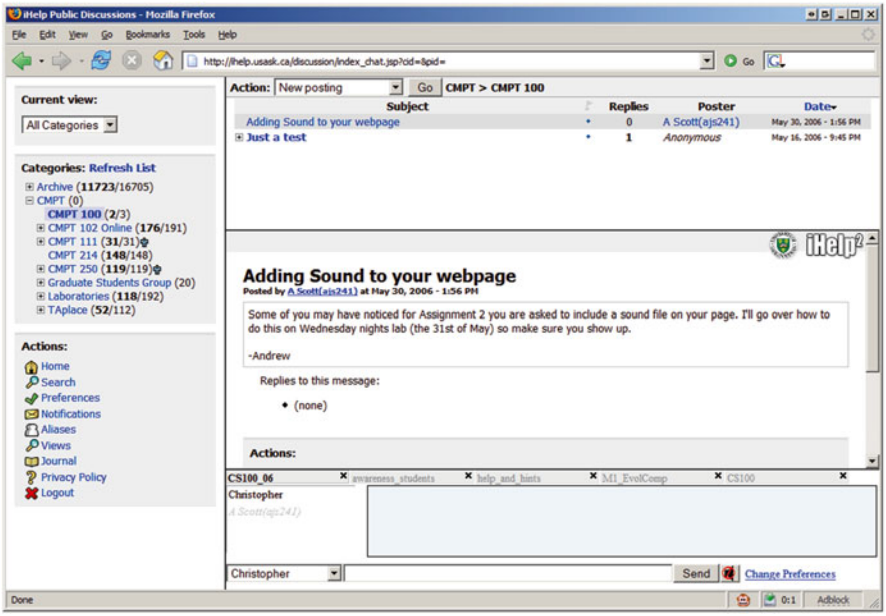

Fig. 7.2 The iHelp Discussion forum system c.2006. The *left side* shows the different forums that a user has access to, in this case a number of course-related forums as well as social and employment forums. Forums are heirarchical, and may include sub-forums related to special topics. Each forum is made up of heirarchical *threads* of discussion as shown in the *top window*. Content for a given *posting* in a thread is shown in the *middle* of the application, and other e-learning tools (such as a synchronous chat system) can be shown at the *bottom* of the application

7.3.2.1 The iHelp Discussions Environment

The iHelp Discussions learning environment (Fig. 7.2) is a Web 2.0 asynchronous discussion forum intended for use within higher education. It was developed out of the iHelp research project (Greer et al. 2001) with the goal of increasing usability and scalability of a technology-enhanced learning environment supporting peer help. The system was deployed within the Department of Computer Science from 2004 to 2010, and was used by thousands of students in dozens of courses annually. When it was retired from general use by the Department in 2010, it contained over 75,000 messages in over 2,200 forums created by over 3,300 users.

This environment is different from other web discussion forums in that it has been augmented to record learner interactions at a fine-grained level. Where other systems typically would show all of the messages within a thread at one time, iHelp Discussions has nested lists of messages which allow the system to record when a user requests a message and how long the user stays with the message open. Further, the forums are hierarchical in nature, and access to forums is logged using similar mechanisms. Over the 6 years it was deployed, more than 3,000,000 read requests were issued by users, with the top message being read over 1,700 times.

The use of iHelp Discussions as a forum ranged broadly by instructor and course. It was commonly used in large cohort undergraduate courses, and there were several department-wide “off topic” forums that students used to discuss issues of technology, philosophy, and politics (amongst others). Access to the forums was restricted through an institutional username and passwords, and public access to messages was not available. In some forums learners were permitted to post anonymously, though the back-end system is still able to disambiguate usernames if needed.

Instructors were free to structure sub-forums however they wanted, and it was common for topic-based, course structure-based, or a flat single forum environment to be used.

One unique aspect of the iHelp Discussions deployment situation is how access is granted to instructional experts. It was not uncommon for instructional experts such as instructors and tutorial/lab assistants to be routinely given access to all of the undergraduate forums. This, along with the detailed usage tracking, makes it possible to study the behaviour of instructional experts as well as learners.

7.3.2.2 Visualization of Learner Activities

Sociograms are a common method of visualizing asynchronous discussions, and have been used by a number of researchers to visualize email correspondence in particular (e.g. Weskamp 2003). These visualizations are graph-based structures where nodes represent individuals in the community and arcs between nodes represent the creation of replies to a message. Nodes are typically rendered using different sizes to represent the status of an individual in the community or the amount of discussion the individual has contributed (e.g. Weskamp 2003). Nodes can be arranged in a number of ways: a strict hierarchy which outlines the abilities or status of groups of nodes is common, as is a physics-based “force graph” which moves nodes closer to one another depending on the characteristics they share.

A representation for the iHelp Discussion forums was formulated using sociograms where nodes indicate particular persons involved in a course and edges between those nodes indicate a discussion replied to relationship. To clearly indicate the difference between learners and instructional experts (e.g. Instructors, Tutorial Assistants, and Markers), each node is colour coded to be either a learner (light grey), or an expert (red). As the iHelp Discussion forums are available to many instructional experts (faculty and assistants assigned to other courses) regardless of the content of the forum, there are typically many red circles in the sociogram.

In large courses (e.g. those with more than 100 students) this formulation quickly became unwieldy. To address this, individuals are further broken up into membership in one of three categories:

- Participants: Those individuals who have written messages, either on their own or as replies to other messages.
- Lurkers: Those individuals who have read postings but have not written any.
- Non-users: Those individuals who have never read nor written a posting.

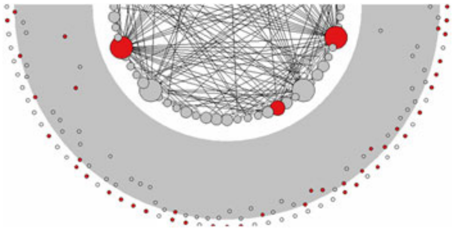

Fig. 7.3 Portion of a sociogram from an introductory computer science course that was taught in a blended fashion and had approximately 200 students and contained 254 discussion postings. Darker (*red*) nodes indicate facilitators, while lighter (*grey*) coloured nodes indicate learners. The *inner circle* is made up of participants, four of which are very important to the community (as shown by having a larger node size). A casual observation of this network indicates that, while some learners write a fair bit (many interconnected nodes in the middle), there are lots of learners who haven't ever read anything (the outer ring of non-users), and many lurkers who read very little (as they tend to be closer to the outside of the middle ring instead of the inside of that ring). Note that the ring of non-users includes a disproportionately high number of facilitators as our current deployment gives access to this forum to most staff and faculty in the department (Colour figure online)

Each category of users is put into their own sociogram that aligns nodes along the exterior of a circle. The different sociograms are then layered on top of one another such that the non-users are farthest from the centre of the sociogram, the participants are closest to the centre of the sociogram, and the lurkers are in between (Fig. 7.3). This corresponds well with the perceived participation rate of individuals (*participants* are more central than *lurkers who* in turn are more central than *non-users*), as well as with the sizes of the different categories of individuals (generally there are more non-*users* than there are *lurkers*, and more *lurkers* than there are *participants*).

Lurking is modelled as a continuous variable—one individual can lurk more than another by reading more forum postings. To support this in the visualization, lurking ratios are calculated and node distances are varied from the outer edge of the lurker region to the inner edge, where lurkers who are closer to the inner edge of the sociogram have read more content. This further reinforces the idea that users who are central in the overall visualization are participating more in the course than learners that are close to the edge of the screen.

Initial feedback from instructors indicated that the act of posting a message does not mean that a user has contributed in a meaningful way. To represent a measure of *importance* in a course, the size of an individual node is varied by the number of persons who have read a user's postings. The calculation for an individual's *importance* is given in Eq. 7.1.

$$\text{Importance} = \frac{\text{Number of people who read my postings}}{(\text{Number of participants} + \text{Number of lurkers}) \times \text{Number of postings}} \tag{7.1}$$

A number of observations about the sociograms can be made based on informal interactions with instructors who reviewed the prototype. In particular, it was observed that:

- A highly connected graph indicates that learners are communicating with one another, while a graph where many nodes are connected only to an instructional expert indicate little peer collaboration.
- The degree of edges coming out of expert (red) nodes indicates how much direct control an instructor has over conversations. Instructors who wait to answer questions have very few arrow heads pointing at their node, while instructors who provoke discussion have many arrow heads point at their node.
- Lurking rates are highly variable, and the majority of lurkers read fewer than 30 % of the postings. This includes non-learners (e.g. tutorial assistants or other instructors), a result that surprised many of the instructors who saw the visualizations.

The next four sections provide specific examples of how this visualization has been used by instructors, and demonstrate the effect of making learner traces more readily available.

7.3.2.3 Example One: Online Small Cohort Course

Many instructors change their pedagogical approach when teaching in online environments, in order to match the needs of learners. In these situations, learners can more easily lose a sense of shared community or shared purpose as the activities and actions of their classmates may be hidden from them. The instructor feels this change too, and many of the consequential awareness indicators that they would normally get from an in-person teaching environment (e.g. attendance and interaction in the lecture, or students coming to tutorials and office hours for assistance) are missing. By revealing indicators of activity in discussion forum systems, the data-assisted approach can help instructors to understand what the virtual classroom interaction looks like.

The Department of Computer Science offered an introductory course for non-majors on the topic of basic Computer Science history, principles, and techniques. In 2006 this course was taught simultaneously to a large cohort of over 100 learners in a blended manner, and to a small online group of 20 learners who had no face-to-face instruction. While taught by different instructors, the online instructor was also the creator of content, assignments, and examinations for both sections.

The online instructor was a firm believer in using the iHelp Discussion forums to build a sense of community amongst the students. Especially in the online course, she saw the discussion forums as the main method of engaging with learners. To this end, one of the weekly requirements was to write an online forum posting about a prescribed course topic and this requirement was mandatory for the online learners only. She would post the initial message indicating what the weekly topic was, and learners were expected to respond with details of content they found on the Internet.

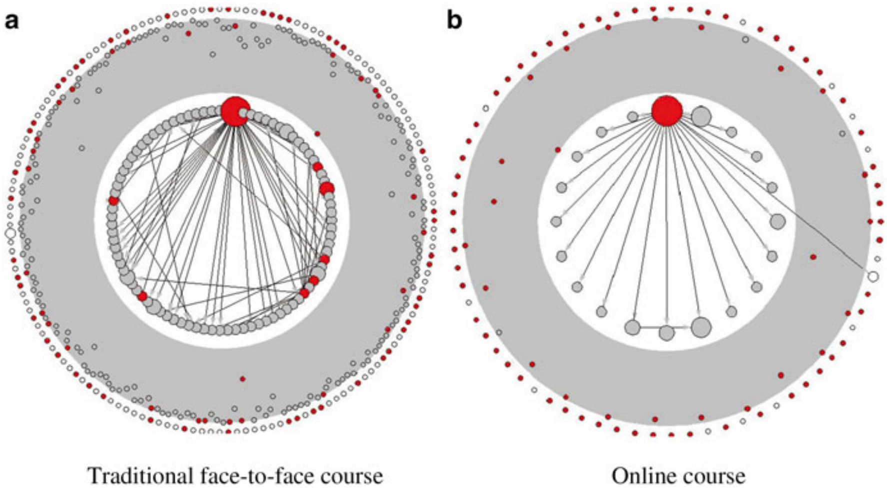

Fig. 7.4 Visualizations of an introductory Computer Science course for non-majors. (**a**) Shows interactions amongst learners who attend traditional lectures, while (**b**) shows interactions amongst learners in a completely online environment. In the traditional course there are many lurkers who have read only a few messages, while the online course required participation of learners and thus has very few lurkers (only other instructional experts). In both cases the instructor is the principal actor in the social network (*large dark* (*red*) node), and learners rarely reply to anyone but the instructor (Colour figure online)

Her reasoning was that by sharing the results of this activity publicly, learners would form a sense of community with one another. A number of weeks into the course, the instructor was shown visualizations for both the in-class and online discussion forums (Fig. 7.4).

The visualization was explained to her, and she showed particular interest to the way size of nodes was generated. She was bothered by the fact that she could easily identify herself as the large dark (red) node who was connected to most students. She wanted students to communicate and form a community with one another, and not just herself. That the pattern existed in both the online and the traditional cohort didn't surprise her, but she felt that traditional learners had other mechanisms by which they form a shared sense of community and thought that making the assignment mandatory for the online learners would be enough to encourage broader use of the discussion forums. In a subsequent offering of the course, the instructor changed the weekly assignments to be more problem-solving based, and set the evaluation criteria such that learners would interact more. The result was a discussion graph that was more fully connected and where learner nodes varied in size (thus *importance*) relative to the instructor (Fig. 7.5).

The use of the data-assisted approach provided insight that was otherwise lost to the instructor. By visualizing the hidden traces learners leave, the instructor became able to understand the effect of her pedagogy, and make interventions that were reflected through subsequent visualizations. Further, the instructor was

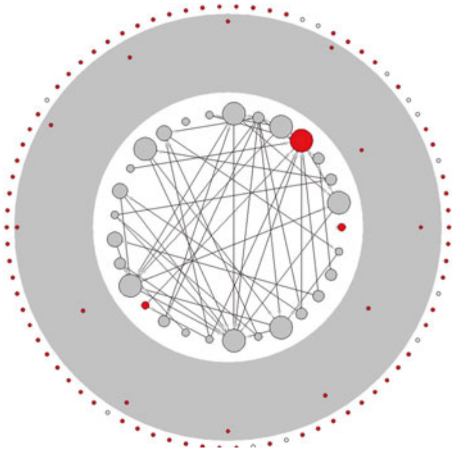

Fig. 7.5 A visualization of a subsequent offering of an online introductory Computer Science course for non-majors. This course was offered after the instructor made pedagogy changes aimed at having learners interact more with one another. In this example, there are several learners who are *important* to the community as shown by their large node size and high level of connectedness

able to compare the familiar traditional face-to-face cohort to a new mode of teaching, and create the kind of learning environment she thought was important for online learners.

7.3.2.4 Example Two: Face-to-Face Large Cohort Course

In traditional face-to-face courses, instructors have the ability to make direct observations of learners. But these observations are minimal in bandwidth, especially when dealing with large cohorts, and tend to only give instructors a broad understanding of the issues learners face. Especially in introductory courses where many of the learners are new to both the discipline and higher education, there is often a hesitancy to speak up in the classroom. The hidden traces of online activities can be leveraged in these circumstances if the traditional course is augmented with technology and taught in a blended mode. Blending a course generally changes the lectures minimally, but offers other avenues for help and exploration (such as discussion forums).

A visualization (Fig. 7.6) of a large cohort of learners involved in introductory Computer Science courses for majors was shown to one of the instructors who teaches in a blended fashion. Unlike the study described in Sect. 7.3.2.3, all learners taking this course (regardless of section) were shown one discussion forum. The instructor shown the visualization was a lecturer as well as the overall administrative organizer for the course, and he would often leave the discussion forums open throughout the day so he could answer student questions. He immediately identified the nodes that represented himself and one of his tutorial assistants (large dark (red) nodes), and theorized on the identities of several of the learner nodes using the size and in-degree. He seemed comfortable with the visualization as an interpretation of the community formed in his course, and felt that the results he had achieved were those he set out to achieve.

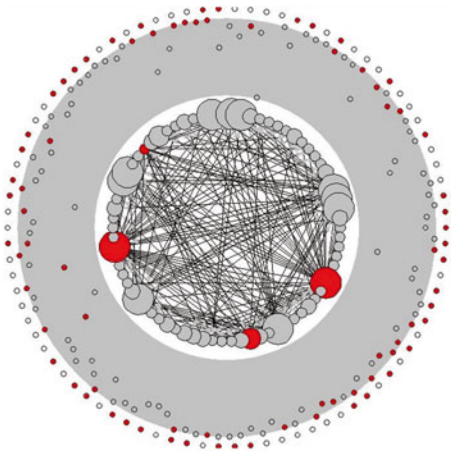

Fig. 7.6 A visualization of an introductory course in Computer Science for majors. This visualization is made up of several sections of learners, each with their own instructor. The high degree of connectedness along with large student nodes (*light grey* in colour) supported the instructor's notion that course discussion would be sustainable if he reduced his level of involvement

A follow-up interview with the instructor provided surprising results; despite being pleased with the interpretation that resulted from the visualization of his course, he still modified his pedagogical approach. He indicated that for the first portion of his course he would typically answer questions as soon as they appeared in the discussion forums. He knew from the volume of questions being asked that learners used the system heavily, and he wanted to address those questions publicly as fast as possible in order to minimize their waiting. However, after seeing several large learner nodes in the visualization (e.g. several learners were read regularly by their peers), he felt that the community had reached a level where peer help would be sustainable without his intervention. In short, he felt he could reduce the speed at which he replied to learners without negatively affecting their learning experience, as their peers were active both in writing responses to questions and in having those responses read by classmates. He was shown a follow-up image of the social network for his course after some time, and remained satisfied that interactions were still happening at a good pace.

In this case, the data-assisted approach was used to confirm a belief an instructor held which would have been otherwise untestable. By aggregating the low-level read events captured by the forum system into a simple metric (*importance*), a small feature of the visualization became enough to convince the instructor of the self-sustaining nature of his course discussions.

7.3.2.5 Example Three: Comparison of Visualizations based on Granularity

The heirarchical nature of the iHelp Discussion forum system allows instructors to create niche topics for their courses. One such example of this comes from an introductory course in Computer Science for majors, where most of the discussion

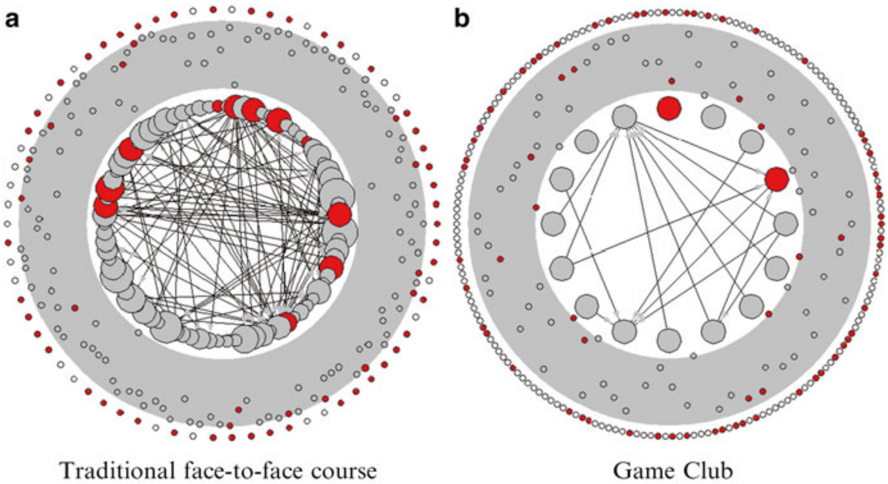

Fig. 7.7 Visualization of a general discussion forum (**a**) and a more specialized discussion forum which is unrelated to course content (**b**). The user cohort is the same, though qualitative remarks about participation in and importance of the discussions can be made by comparing visual aspects of the two renderings

happened in a general discussion forum (726 of 901 messages), but some happened in topic-specific forums. This course taught general principles of Computer Science, but based on the interests of the instructor and the students a sub-forum called the *Game Club* was created to talk about video game-related issues. These two discussion forums can be visualized individually (Fig. 7.7) and, while no discussion with the instructors about this forum took place, some high level qualitative comments can be made.

First, it is easy to see that there is much more activity in the general discussion forum than the game forum by looking at the in-degree of participants (higher in general forum) and the number of persons in the non-user sociogram (higher in game forum). Second, the rate of lurking is higher in the more specialized game forum, though the cause of this is unknown (it might be that there are just fewer postings, so it is easier for the keen learners to read them all, or there is an initial surge of activity where the community has promise but then dies out). Finally, it is worth noting that a number of instructional experts sit very close to the participation ring in the game forum, but do not write new messages or reply to messages to become participants. The precise role of the instructional experts here is not clear and it is difficult to draw specific conclusions; they may be instructors who are keeping an eye on developments, or tutorial assistants who are peer students and very much interested in the content of the discussions.

Insight coming out of use of the data-assisted approach is very much about making visible the invisible—instructional experts cannot regularly see interactions learners make and thus do not involve them when making pedagogical decisions.

In this example, an end-user looking at discussion forum postings without the visualization would assume that only the two instructional experts who have written messages in the game forum are interested in the topic. The visualization, however, makes it clear that this is not the case, in that there are eight other instructional experts who have read almost every message posted in that forum. By making visible the hidden traces users leave as they interact with the learning environment, instructional experts are better positioned to make both short-term and long-term pedagogical decisions. Given the keenness of learners to view this game-related forum, it might not be unreasonable for similar content to find its way into the standard curriculum for the course in future offerings.

7.3.3 Case Study Two: Measuring Educational Impact

As shown in the previous section, the data-assisted approach can be used to generate insight which instructors can use to modify their teaching activity, and each of the examples given describes how broad changes in pedagogy that could be made based on understanding the hidden behaviours of learners. But the data-assisted approach is not limited to just instructors or broad changes; it is possible for instructional designers, for instance, to leverage similar techniques to make changes that affect only a portion of the learner population. This section[2] describes how learner behaviour data can be statistically described and related directly to educational outcomes. It further provides methods to model learners and associate them with pedagogically sound groups where more individualized interactions can take place. In doing this, this section addresses three questions that come from the motivational scenario in Sect. 7.2.3.2:

- Can learners be clustered based on their viewing habits into pedagogically relevant groups?
- If so, do these groups differ in their formal assessment measure?

7.3.3.1 The Recollect Environment

A key consideration of the data-assisted approach is whether learner interactions within the learning environment can be correlated with pedagogical goals and measures of learning outcomes. These interactions can be either explicitly made by learners (e.g. through the filling out of a survey) or implicitly made as a by-product of the learning activity itself (e.g. navigating through content). Sometimes referred to as *clickstream* data or *traces*, these interactions are difficult to understand on their own in part because of the large amount of data collected (potentially millions of data points) and the low level meaning that the data represents (e.g. the clicking of

[2] Portions of this section appear in Brooks et al. (2011a).

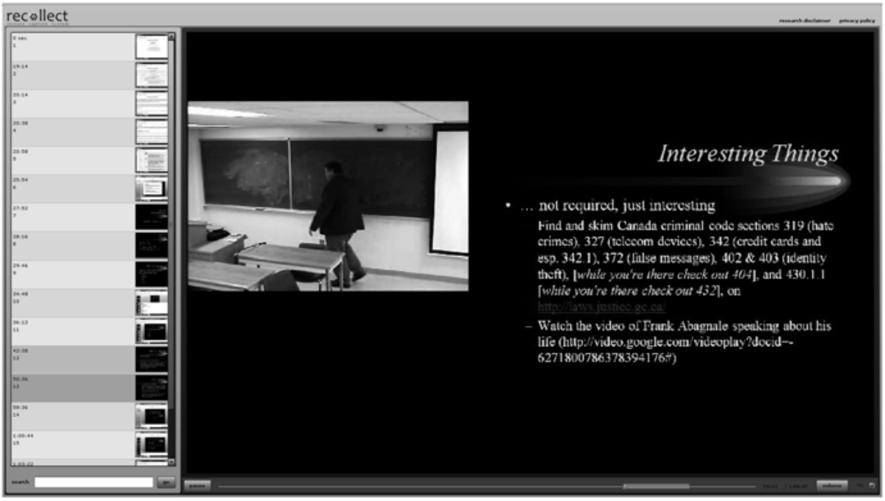

Fig. 7.8 A screenshot of the Recollect lecture capture system. The system shows thumbnails of upcoming slides on the *left* of the window to allow for navigation. The *right side* of the player shows video of the classroom as well as video captured from the data projector

a single link or keystroke on the keyboard). The data must first be summarized, and then linked to learner goals in order to be made actionable.

One learning environment that collects this low-level learner behaviour data is the Recollect lecture capture solution (Fig. 7.8), developed at the University of Saskatchewan in part by the authors. This system records in-person classroom lectures and stores them for playback by students for both the initial viewing of and reviewing of content. Recollect records a number of user behaviours, including the time spent streaming a lecture (discretized into 30 s intervals), clicks on any buttons in the user interface (e.g. volume change), searching through lecture slide content, seeking within the video using the video scrubber, and navigating within the video using section thumbnails. Each of these behaviours is linked to the time in which they were observed, the student who initiated the behaviour, and the particular video that was being watched.

7.3.3.2 Formal Assessment

The Recollect system was deployed for a number of sections of a second year Chemistry course in both the 2010 and 2011 academic years. Students were allowed to use the system how they saw fit, and every lecture from a single section taught by one professor was shared with students in all sections of the course. Instructors did not change grading criteria based on the presence of the recorded lectures, and midterm and final examinations were common across all sections of the course.

Lecture capture is only one resource learners had available to them, and consistent patterns are difficult to see from the raw data. The viewing behaviours for Chemistry 2010 learners were cleaned[3] and summarized into viewing habits broken down by calendar week. In this model, learners were deemed to have either watched or not watched lecture content during the 12 weeks of the course. Out of 636 learners registered in the course, 133 were included in the study (participation rate of 36.6 %) by virtue of their use of the Recollect system. The weekly viewing rates of learners were used as attributes with k-means clustering ($k=5$) to create a model of learner behaviours. Five clusters were chosen based on preconceived hypotheses of how learners might use the system (see Brooks et al. 2011a for more details). The results of the clustering activity provide a number of insights into learner activities. For instance, some learners only watch lectures during the week of the midterm, while others watch fairly regularly. Regardless of viewing patterns, the last 2 weeks of the course (corresponding to the time between the end of classes and the final examination) tended to have a high amount of disagreement between participants and the centroids. The disagreement for these weeks, ranging from 19 to 40 %, suggests that activity throughout the term isn't indicative of behaviours between the end of term and the final exam, and thus only data during the teaching portion of the term was used for further analysis.

Using the data provided from the initial 2010 students, a high level model for five idealized clusters was developed. In this model, first cluster has learners who habitually watch lectures throughout the term (*high activity learners*), the second cluster is made up of learners who observed the lecture the week before the midterm examination (*just-in-time learners*), the third and fourth cluster appear to correspond to (roughly), the first and second half of the course (*disillusioned learners* and *deferred learners*) respectively, and the last cluster is made up of learners who did not watch many lectures, though they must have watched at least 5 min of video in a week to be included in the study (*minimal activity learners*).

With this high level model defined, a learner from any cohort can be placed into a particular group based on similarity. For instance, a learner who watches video every week except for the first and sixth weeks will be placed into the *high activity* cluster. Despite this learner not fitting perfectly with this cluster, his or her activity patterns are most closely related to it. Thus the centroids are not the only interesting aspects of the clusters, the amount of error is as well.

Instructional goals are often represented by midterm and final examinations as a proxy for learning. Correlating patterns of behaviours with differences in grades provides some evidence of learning from activity. Learners use lecture capture as one tool to aid in learning, but many other tools and methods contribute to learning (e.g. online quizzes, in-class lectures, textbooks, study groups) and make identifying the effect of any single tool difficult. A pairwise tukey test for the midterm

[3] A threshold of at least 5 min of viewing was arbitrarily chosen to remove behaviours that were deemed to be tool experimentation over tool use for learning. As the time period for this course was in the second semester of the academic year, the 1 week of data over midterm break was excluded from analysis.

Table 7.1 Midterm, final examination, and overall grade averages and standard deviations broken down by cluster in percentages for the Chemistry 2011 course

Cluster label	n	Midterm$_{\bar{\chi}}$	Midterm$_{\sigma}$	Final$_{\bar{\chi}}$	Final$_{\sigma}$	Overall$_{\bar{\chi}}$	Overall$_{\sigma}$
High activity	14	77.32	15.71	75.43	22.19	80.14	14.49
Disillusioned	18	64.71	15.72	58.98	17.81	66.82	13.85
Just-in-time	86	68.14	15.92	60.49	23.68	70.69	14.39
Minimal activity	191	64.30	15.36	58.98	22.89	68.41	14.71
Deferred	24	63.33	12.20	59.83	20.21	69.04	12.43

examination, final examination, and overall marks for the initial cohort demonstrated that there is an effect on marks for one cluster of learners in particular, the *high activity learners*, and that the effect's significance ranges between the levels of $p=0.021$ and $p=0.240$. Table 7.1 shows the difference in grades between the different learner clusters. Not shown in this table are the average incoming GPA values of each cluster, for which there was no statistically significant difference.

7.3.3.3 Learner Goals

Formal assessment is not the only indicator of learning, and a similar approach using machine learning allows one to form a relationship between the behaviours of learners and subjective questionnaire data. This can be useful in many ways—for instance, if a relationship is discovered between low use of lecture capture and negative opinions of the technology, the instructional expert may be able to change the learning environment in the future to accommodate learners who would prefer alternative tools based on their activity alone.

Armed with knowledge of the domain and pedagogy, domain experts like Michelle can query learning environments to gain a deeper understanding of how learners are acting within groups. Such queries are likely to be driven by hypotheses based on curiosity, preconceptions based on training, and instincts based on years of practice. To emulate this investigation, seven questions about the usefulness of the system and perceived workload were examined with respect to how well they fit clusters[4] ($k=2$) based on two behaviours: The number of minutes the learner watched and the number of unique videos the learner watched. The goal in doing this was to see if activity could be linked to statistically significant differences in learner opinions.

Learners ($n=636$) in the Chemistry 2011 cohort were surveyed as to the relevance and usefulness of the Recollect system in this class. The questions asked covered a mixture of technical, pedagogical, and policy issues, and a number of these questions were designed to elicit beliefs learners had about their learning (response rate of $n=229$, 30 %). The full survey instrument can be found in Brooks

[4] The choice of the number of clusters (i.e. the value of k) to make affects outcomes greatly. This was an initial investigation to determine if unsupervised machine learning approaches can be used for clustering of subjective responses to data. Given the results shown here it is reasonable to continue exploration with an aim to find ideal values for k.

Table 7.2 Student behaviour clusters based on the number of minutes watched and unique videos watched with $k=2$

	Clusters				
	Keen ($n=12$)		Less keen ($n=115$)		
Attribute	$\bar{x}$	σ	$\bar{x}$	σ	ANOVA p
Minutes watched	1143.25	414.37	187.30	236.51	≤0.001
Unique videos	28.25	7.10	6.93	6.43	≤0.001
q4	1	1.04	1.16	1.01	0.61
q5	1.5	1.31	1.70	1.08	0.56
q6	0.59	0.80	0.79	0.77	0.04
q9	0.42	0.51	1.19	1.03	0.01
q21	0.25	0.87	0.44	0.97	0.51
q22	0.17	0.39	0.99	1.30	0.03
q24	0.50	0.80	1.25	1.28	0.05
Midterm	72.08	18.49	71.33	15.14	0.87
Final	72.22	21.15	70.41	20.00	0.77
Overall	77	15.32	75.75	13.15	0.76

Cluster labels, *keen* and *less keen* were added by the author as descriptive elements only, and are arbitrary. The only strong correlation with questionnaire data existed for questions 6, 9, 22, and 24

(2012). The seven questions looked at were all based on a five-point scale, with the first two being an increasing scale of marks, and the last five being a likert-based scale where zero represented the strong affirmative (e.g. “Very High” or “Very Important”) and four representing the strong negative (e.g. “Very Low” or “Not Important”). The question text was:

- Q4: I am working in this class to try and get a mark in range:
- Q5: Reflecting on my performance in the class so far, I think my mark will actually be in the range:
- Q6: My workload this term including all of the courses I am in as well as other commitments is:
- Q9: How important do you feel that watching the recording of the lecture was for your success in this class?
- Q21: If you used the lecture capture system, how important was it for reviewing content you hadn’t seen (e.g. missed classes).
- Q22: If you used the lecture capture system, how important was it for reviewing content you saw but didn’t understand or couldn’t remember?
- Q24: If you used the lecture capture system, how important was it for studying for examinations?

The analysis of these questions with the clusters formed is shown in Table 7.2. Labels on clusters were chosen by the authors, and learners were segmented into a smaller cluster ($n=12$) of users who watched a large number of videos (on average, 28) for a mean time of 19 h and 3 min. A larger number of learners ($n=115$) watched fewer videos (on average, 6) for a mean time of 3 h and 56 min. Only questions 6, 9, 22, and 24 showed statistically significant results ($p \leq 0.05$), though the means between clusters for question 6 were of little meaningful difference.

7.3.4 Case Study Three: Adapting Learning Environments to Tasks

The scenario presented in Sect. 7.2.3.3 followed an educational technologist, Adam, who used the data-assisted approach to understand how learners are using the lecture recording tools. More than just gaining *insight*, Adam was interested in building *instructional interventions* of an automated manner. Using clustering techniques like those described in the previous chapter, it is possible for an instructional expert like Adam to identify interesting groups of learners, and create an intervention. Thus far, however, only broad pedagogical interventions executed by instructors or instructional designers have been described. One of the interesting aspects of traditional intelligent learning environments (such as ITS) is that they respond automatically to learner actions. In this section[5] we consider whether this ability is lost in a data-assisted approach, where instructional experts are expected to be involved in the sensemaking process. In particular, we look at two questions:

- Do groups of learners really agree on where indices should be placed, or are their preferences for navigational aids more varied? If the former is true, then the clustering methods described previously may well yield a more personalized and efficient navigation structure.
- Is it appropriate to use supervised machine learning to build indices, and how might such an approach compare to algorithms that already exist?

7.3.4.1 Navigation in Recollect

The Recollect lecture capture environment (Fig. 7.9) has multiple methods a learner can employ to navigate through content. For instance, thumbnails across the left hand side of the environment allow for quick "chaptering" of the content with image preview, while the scrubber along the bottom allows for precise navigation throughout the video based on time. It is not unreasonable to think that navigational style might differ depending on learning goal; a learner watching a lecture for the first time might not use either of these navigational aids; a learner who is searching for a particular topic in the lecture might use the thumbnails provided; a third kind of learner might use the scrubber to quickly replay video about a critical concept they missed while watching.

Thumbnails in the Recollect environment were originally generated using a naive algorithm based on Time—every 5 min of video a still image would be copied from the video and metadata for the video would be updated linking the image and its position in the video. More sophisticated methods have been proposed for the same purpose; for instance, the Opencast Matterhorn system (Brooks et al. 2011b) uses a frame differencing algorithm with thresholds for RGB colour values, while

[5] Portions of this section appear in Brooks and Amundson (2009).

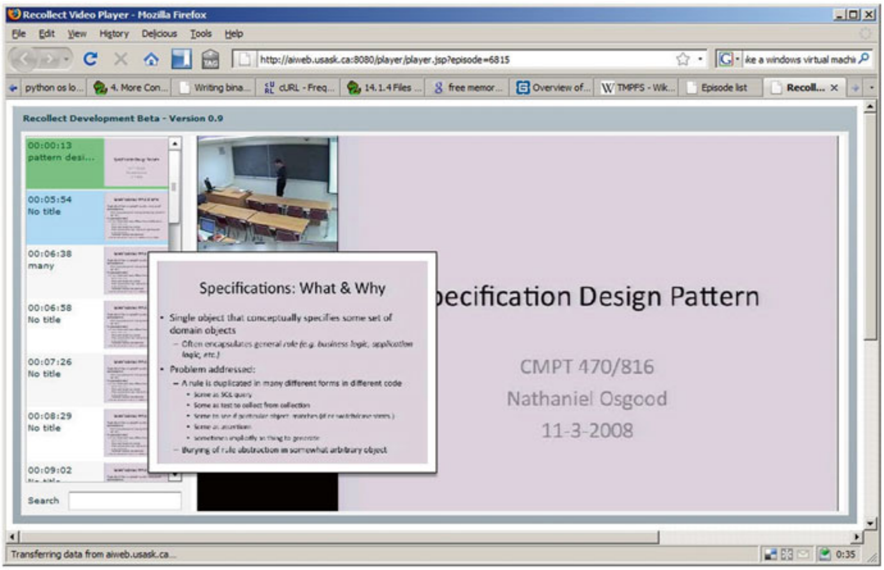

Fig. 7.9 The Recollect lecture capture system, showing navigational thumbnails on the *left hand side*. As users mouse over a given thumbnail, a small image opens up and shows what the data projector feed recorded at the corresponding time in the video. All interactions involving the thumbnails such as mousing over, clicking, or scrolling through the list, are recorded

Dickson's algorithm (Dickson et al. 2006) is a multi-pass image processing function that examines both pixel and block characteristics of video to determine stable events. Both of these algorithms were designed to work with lecture video captured by similar hardware as used by the Recollect system, making the potential for comparative study possible.

7.3.4.2 Comparing Users Actions to Traditional Algorithms

A laboratory study was undertaken to determine whether indices of video could be created for lecture video based on learner opinions of relevance. Indices are used by Recollect as thumbnails for navigating through a lecture, as shown in Fig. 7.9, and correspond roughly with DVD chaptering. If successful, such a method may be appropriate for generating indices in an ad hoc manner—an important result when applying the data-assisted approach.

Six human subjects who were unfamiliar with lecture capture systems were asked to go through four different lectures and identify where significant events occurred. The tool provided to study subjects allowed for navigating through a video linearly both forwards and backwards in one, five, 10, and 30 frame increments (each frame was equivalent to 1 s of video).

Observations and survey results from the participants identified that they used two distinct mechanisms for identifying significant events: visual structure (e.g. slide advancement in PowerPoint, or extending the canvas in the Sympodium) and semantics (topics being taught in the slides). Out of the six participants, five used primarily visual structure to identify events, while the sixth used the semantics of the lecture material. The level of agreement between participants excluding this sixth rater ranged from $\kappa = 0.18$ (slight, according to Landis and Koch 1977) to $\kappa = 0.87$ (almost perfect, according to Landis and Koch 1977). The videos that held traditional PowerPoint slides all had high inter-rater reliability ($\kappa \geq 0.66$).

This suggests that the previous methods of unsupervised machine learning we have discussed may be appropriate for generating groups of learners who navigate through lecture video similarly. This is useful for an educational technologist like Adam, who is seeking to modify the learning environment in order to improve indexing. Subsequent investigation as to the performance of indexing algorithms like Time, Opencast, and Dicksons's indicated they only poorly matched human raters. Instead, we look to leverage that rating data with supervised machine learning in order to personalize the indexing method. In a real-world environment, this rating data might come from logging data or social bookmarking behaviour of learners.

7.3.4.3 Adapting Navigation Based on Supervised Machine Learning

Most thumbnailing methods approach the issue of forming indices in lecture video as an image recognition problem. The goal of these methods is to measure the difference between two or more frames of video, and use this with some threshold value to determine when a significant change has occurred. The problem with this approach is in the selection of and weighting of attributes that make up the difference function; a data-assisted approach argues that the attributes should not be chosen a priori, but should be customized based on the learner (or cohort of learners) who are using the system.

Supervised machine learning methods take a set of instances, a set of attributes, and a set of classifications and build a model that can be used to predict new classifications for further instances with similar attributes. In the case described here, the set of instances are the video frames shown to subjects, the set of attributes are the image characteristics for these frames which are determined automatically (see Brooks and Amundson 2009; Brooks 2012) and the set of classifications is whether a given image is an index or is not. The output of a supervised method is a set of rules that can be applied to new images to determine if, based on this *training data*, those images are or are not indices.

Having determined that end-users are in a reasonable level of agreement when coming up with video indices, we ran a second study (Brooks et al. 2013) with six new participants in order to collect detailed indexing information. With this data, we formed 6 tenfold cross-validated J48 decisions trees. The trees were formed on modified versions of the training set, adjusting the threshold for the minimum agreement among raters before an instance was considered an index. The thresholds were

Table 7.3 Group κ between raters and algorithms

Comparison algorithms			Our trained algorithms					
Time	*Opencast*	*Dickson*	T_1	T_2	T_3	T_4	T_5	T_6
0.391	0.370	0.448	0.574	0.565	0.565	0.537	0.530	0.487

Upper and *lower* are the max and min values any algorithm could provide for κ. The κ between the six expert raters without an algorithm was 0.577

set such that T_1 was given a positive classification on all instances where at least one rater indicated there should be an index, T_2 is a tree trained on data where the threshold was two raters, and so on, until T_6 required perfect agreement between raters that an instance was an index. The goal was to see what effect different aggregation methods would have on the resulting level of agreement. As shown in 3, the trained algorithms all outperform the traditional methods of indexing, with the most stringent training method (T_6) providing the worst results ($\kappa = 0.487$) and the most lax training method (T_1) providing the best results ($\kappa = 0.574$). Further, each trained algorithm outperformed the static algorithms used for comparison (Table 7.3).

7.3.5 Conclusions

Previously (Sect. 7.2.3), we considered the needs of three different kinds of instructional experts; Katheryn the Instructor, Michelle the Instructional Designer, and Adam the Educational Technologist. Each of these experts is interested in gaining insight into the interactions learners have with technology, and leveraging this insight to create instructional interventions.

In this section, we have described three real-world educational systems that employ the data-assisted approach. In the first of these, human–computer interaction techniques of information visualization were used to aggregate traces of learner activities and make them available to instructional experts. Through augmenting an asynchronous discussion forum, instructors have been able to modify their pedagogical practice and gain insight into how learners interact in niche communities. This demonstrates how applying the data-assisted approach can lead to insights in instructional experts that they can use to modify their teaching practice.

The second system we looked at demonstrated that the traces learners leave behind when using a lecture capture and playback system can be data-mined and related to educational outcomes and goals. Clusters formed the basis for an abstract model, where each cluster represented different learning strategies. One group in particular, the *high activity learners*, correlated well with an increased achievement compared to other groups. With this knowledge, the instructor could apply clustering to future students and build instructional interventions aimed at particular groups. For instance, if it is the instructor's belief that the correlation relationship between regular lecture video watching and higher marks is a causal relationship, he or she might send out an alert to all learners who are not watching videos to encourage them to watch more consistently.

Continuing with a look at the Recollect environment, the third case study used a mixture of qualitative and quantitative laboratory studies to create a method for adapting the presentation of navigational indices in the user interface of a lecture capture system. By combining learner opinions of significance with supervised machine learning techniques, we demonstrated that substantially higher levels of accuracy of navigational indices can be achieved. Such a result demonstrates that there is value in using the data collected from learning environments to change the environment itself through instructional experts.

7.4 Discussion and Conclusions

Students learn better in more individualized tutoring situations (Bloom 1984), a result that has spawned two decades of intensive research in intelligent learning environments. These environments, such as ITS and adaptive hypermedia systems, deliver content to learners and form models of them based on the a priori definition of pedagogical approaches, learner traits, and content semantics. This allows for personalization of the learning environment which can be realized by changing the content, navigation, or structure of the learning environment for a particular group of learners.

This method of personalizing learning environments is expensive. It requires up-front cost in design and development and, as such, these methods are usually used within a single discipline or course. To scale across different domains, institutions of higher education use simplified learning content management systems. These systems offer instructors a thin technological wrapper around their existing content and learning activities, and provide only minimal support for personalization.

The data-assisted approach presented here supports learning in technology-enhanced learning environments by generating *insight* for instructional experts and enabling this insight to be used for *instructional interventions*. Instead of replacing instructional experts, the data-assisted approach enables them to see the hidden traces learners leave behind as they interact with the learning environments, to understand these traces in light of educational goals, and to apply this insight to form instructional interventions.

7.4.1 Discussion of Findings

There are many different kinds of instructional experts who might use data-assisted approaches: for example, instructors, instructional designers, tutorial assistants, and educational technologists. Each of these groups has different needs. For instance, an instructor might need to be able to understand what problems are faced by a particular cohort of students they are instructing, while an instructional designer or an educational technologist might want to generalize trends across cohorts and obtain insight about particular approaches or tools.

The data-assisted approach is broad enough to address these different cases. Section 7.3.2 described a situation where a variety of different instructors were shown visualizations of student interactions. These instructors taught different courses with different modalities (e.g. online versus blended in Sects. 7.3.2.3 and 7.3.2.4) and different scopes (e.g. communities of interest in Sect. 7.3.2.5). In each of these cases instructors were able to form insights into how learners in their course were interacting, and were able to use this insight to change their teaching practice.

Instructional designers and education researchers are also actors that can engage with data-assisted approaches. Section 7.3.3 describes the use of unsupervised machine learning methods to discover clusters of learners based on their lecture video viewing habits. These clusters correlate well with both pedagogical expectations and educational outcomes. By making visible the hidden viewing habits of learners using a lecture capture system, statements about the efficacy of lecture capture as a study aid can be made. For instance, the evidence that learners who watch lectures regularly have higher outgoing grades suggests that lecture capture may have an impact on learning, an important consideration when designing support for large cohort courses.

The data-assisted approach creates insight with instructional experts through dialogue. In Sect. 7.3.2 this dialogue takes the form of information visualization, and instructors could see different discussion forums in their courses at different times. In Sect. 7.3.3 this dialogue was more interactive, and allows experts to parameterize clustering and select attributes of interest. Regardless, it is the method of explicitly including the instructional expert in the sensemaking process that makes the data-assisted approach suitable for building intelligent educational environments in higher education.

Once insight has been formed, instructional experts need a way to improve learning through instructional interventions. Here again the different roles of experts change the way that instructional interventions are made. For instance, in Sect. 7.3.2 instructors largely developed interventions outside of the technology-based learning environment. One instructor, for instance, changed her assignment requirements which caused learners to interact differently—an interaction pattern she saw as more pedagogically sound. Another instructor used the insight generated from the data-assisted approach to reduce his level of interaction in the class based on a perception that the current discussion environment was already sustainable. In both of these scenarios, the instructor made these interventions based on visualizations resulting from the application of the data-assisted approach.

Instructional interventions are not always broad pedagogical changes, and software systems such as adaptive hypermedia systems often focus on small customizations to the learning environment to improve learner experience for individuals or groups of learners. Section 7.3.4 demonstrated that the data-assisted approach can be used to provide these forms of adaptations as well. Working from data representing ideal indices in lecture video, supervised machine learning was to be used to take prototypes of ideal indexing and apply these to different video content. Many different prototypes for various situations can be formed—those learners who want an overview might get one set of indices, while those who want visual navigation might get another.

Table 7.4 Outline of data, data-processing techniques, and the insight and instructional interventions they might lead to when using the data-assisted approach

Data	Data-processing technique	Insight	Intervention
Student reading data of asynchronous discussion messages	Sociogram-based information visualization	Discover community of practice and level of social engagement	Scaffold discussion (instructor), contribute to communities of interest (instructor)
Student viewing of lecture videos	Clustering of students with *k*-means and statistical treatment of assessment	Identify students with suboptimal study habits	Prompt change in study habits (instructor), recommend lecture video usage (system)
Student navigation in lecture videos	Clustering of students with *k*-means	Discover popular portions of lectures (instructional expert), identify segments of lecture video (system)	Review heavily studied concepts with extra material (instructor). Provide better indexing of video and adaptive navigation (system)

These prototypes do not have to be formed by an expert; instead, they can come directly from the learners themselves, and the expert (in this case an educational technologist) employs insight in designing a technology-enhanced learning environment.

7.4.1.1 Connecting Data to Insights and Insights to Interventions

While in this work we demonstrate several different insights and interventions, it is less clear exactly which data and data-processing techniques lead to these insights and interventions. This question is particularly salient in light of the software engineering task of building personalized learning environments—to form repeatable design patterns ("recipes" for successful software development) that can be used by software developers to build data-assisted software, these developers need to understand what data and data-processing techniques will provide instructional insight.

This issue is multidisciplinary in nature, and requires the consideration of education researchers, human–computer interaction designers, and information retrieval experts. Further, each set of data, data-processing technique, insight, and intervention can be considered at different levels of granularity, making the issue potentially more complex. Table 7.4 provides initial formulations of what such a taxonomy might look like, using the investigations provided in this chapter. The spirit of the data-assisted approach is that the intelligence of the system is the result of a dialogue between software and the instructional expert. In keeping with this, Table 7.4 should be seen as some general guidelines towards design patterns, and not an exhaustive list of which data and techniques lead to specific insights and interventions.

7.4.2 Conclusion

Unlike most other methods of building intelligent learning environments, the data-assisted approach does not seek to replace instructional experts but to actively engage with them. It does this in two ways; by generating *insight* from data through a dialogue with the expert, and supporting experts as they act on this insight to form *instructional interventions*. This allows institutions of higher education to leverage the intellectual support resources they already have (e.g. instructors, instructional designers, and educational technologies) to provide more personalized learning experiences in technology-enhanced environments.

References

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Brooks, C. (2012). *A data-assisted approach to supporting instructional interventions in technology enhanced learning environments*. PhD thesis, University of Saskatchewan, Saskatoon.

Brooks, C., & Amundson, K. (2009). Detecting significant events in lecture video using supervised machine learning. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), *2009 Conference on Artificial Intelligence in Education (AIED09)* (pp. 483–490), Brighton, UK. Amsterdam: IOS Press.

Brooks, C., Epp, C. D., Logan, G., & Greer, J. (2011). The who, what, when, and why of lecture capture. In *Proceedings of the First International Conference on Learning Analytics and Knowledge—LAK'11* (pp. 86–92). New York, NY: ACM Press.

Brooks, C., Greer, J., & Gutwin, C. (2012). Using an instructional expert to mediate the locus of control in adaptive E-learning systems. In *Second International Conference on Learning Analytics and Knowledge 2012 (LAK'12)* (pp. 84–87), Vancouver, BC.

Brooks, C., Johnston, G. S., Thompson, C., & Greer, J. (2013). Detecting and categorizing indices in lecture video using supervised machine learning. In *26th Canadian Conference on Artificial Intelligence*, Regina.

Brooks, C., McKenzie, A., Meyer, D., Moormann, M., Rihtar, M., Rolf, R., et al. (2011). OpenCast Matterhorn 1.1. In *Proceedings of the 19th ACM International Conference on Multimedia—MM'11* (pp. 703–706). New York, NY: ACM Press.

Brooks, C., Winter, M., Greer, J., & McCalla, G. (2004). The massive user modelling system (MUMS). In *Intelligent tutoring systems 2004 (ITS04)* (pp. 73–112). Maceio-Alagoas, Brazil: Springer.

Buskist, W. (2004). Ways of the master teacher. Association for Psychological Science: Observer. Available online at http://www.psychologicalscience.org/index.php/uncategorized/ways-of-the-master-teacher.html.

Carnegie Learning. (1998). *The cognitive tutor: Applying cognitive science to education. Technical report*. Pittsburgh, PA: Carnegie Learning, Inc.

Dickson, P., Adrion, W., & Hanson, A. (2006). Automatic capture of significant points in a computer based presentation. In *Proceedings of the eighth IEEE international symposium on multimedia* (pp. 921–926). San Diego, CA: IEEE Computer Society.

Greer, J., McCalla, G., & Vassileva, J. (2001). *Lessons learned in deploying a multi-agent learning support system: The I-Help experience*.

Klein, G., Moon, B., Hoffman, R. R., & Associates, K. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88–92.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

McCalla, G. (2004). The ecological approach to the design of e-learning environments: Purpose-based capture and use of information about learners. *Journal of Interactive Media in Education, 2004*(May), 1–23.

Najjar, J., & Wolpers, M. (2006). Attention metadata: Collection and management. In *Workshop on logging traces of web activity: The mechanics of data collection*, Edinburgh, UK.

Peckham, T., & McCalla, G. (2012). Mining student behavior patterns in reading comprehension tasks. In *Proceedings of the Fifth International Conference on Educational Data Mining* (pp. 87–94), Chania, Greece.

Siemens, G. (2012). Sensemaking: Beyond analytics as a technical activity.

Suraweera, P., & Mitrovic, A. (2002). KERMIT: A constraint-based tutor for database modeling. In *Intelligent Tutoring Systems: Sixth International Conference, ITS 2002* (pp. 2012–216), 2–7 June, 2002. Biarritz, France: Springer.

Weskamp, M. (2003). *Social circles*. Retrieved March 2006, from http://www.marumushi.com/apps/socialcircle

Chapter 8
Identifying Points for Pedagogical Intervention Based on Student Writing: Two Case Studies for the "Point of Originality"

Brandon White and Johann Ari Larusson

8.1 Introduction

One of the trends in the American university system over the past several years has been the migration towards larger and larger so-called gateway courses (MacGregor 2000), which are typically a student's first exposure to collegiate work. These courses, with familiar names like Economics 100 or Biology 101, are classified as "gateways" because they are the single portal through which students must pass to gain entry to more advanced courses. Large numbers of students are enrolled in these courses, and, consequently, instructors have less opportunity than normal to keep track of how well or how poorly individual students are coming to understand the course material. Unfortunately, these large classes have a pronounced impact on students' learning process.

Large lectures are useful for conveying blocks of information, but less useful for fostering the kind of higher-order thinking (McKeachie and Chism 1986; Cooper and Robinson 2000), that would be of the most use later in a student's academic career.

For most if not all learning activities, a substantial amount of an instructor's time and effort is devoted to evaluating and monitoring the quality of students' work, and thus, hopefully, the depth of their learning (Crooks 1988). The purpose of this monitoring, however, is not merely the determination of grades; part of the instructor's work is entirely self-reflective, enabling the instructor to concurrently, or ideally even preemptively, intervene to make adjustments to course pedagogy based on students' engagement or understanding (McAlpine et al. 2004).

B. White (✉)
Department of English, University of California, Berkeley, Berkeley, California, USA
e-mail: brandonw@berkeley.edu

J.A. Larusson
Center for Digital Data, Analytics and Adaptive Learning at Pearson,
Boston, Massachusetts, USA
e-mail: johann.larusson@pearson.com

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7_8, © Springer Science+Business Media New York 2014

Today it is not uncommon for faculty to deploy an online instructional environment in order to relieve some of the possible consequences of large classes. Technological solutions, like producing discussion posts or blogging about the material, can be used to make the class seem smaller, promoting student conversations about course materials and ideally heightening comprehension and retention of course readings (Larusson and Alterman 2009). One positive by-product of the use of activities like blogging is that it permits an instructor to see each blog post or each discrete writing sample as a window into a student's comprehension. With writings produced regularly, on an iterative basis, instructors have several possible points at which a student's progress might be evaluated even before a graded assignment comes due.

In a class with hundreds of students, however, this task becomes more difficult if not impossible. Manually determining which students have improved or which have regressed over the span of many weeks, would prove an intensely laborious process, far more complicated than simple reading and re-reading of a single student's work. For a technologically mediated course, however, the problem becomes somewhat (if only slightly) more palatable; with work produced in digital form, it becomes possible to create automated and electronic instructional aids (Greer and Heaney 2004) that can help an instructor identify just what is taking place.

This chapter describes an automated solution that can be used by educators to help resolve these tensions. While the electronic signature of online writing samples have the potential to serve as evidence of students' progress, it is perhaps more difficult to characterize the quality of the learning taking place. Through the application of lexical analysis to student writing, we have developed and tested an analysis method and tool that allows an instructor to track how a student's written language migrates from mere paraphrase to mastery, isolating the moment when the student's understanding of core concepts best demonstrates an ability to place that concept into his or her own words, a moment that we've chosen to call the "Point of Originality." This process recreates the same cognitive activity that educators might ordinarily undergo, yet in an automatic manner that is dramatically less labor intensive. Ultimately, the resulting data is presented to the instructor by way of a custom visualization, which allows for continuous self-monitoring with minimally expended effort.

Such an analysis method and tool can be understood as part of the burgeoning field of learning analytics. Learning analytics are broadly defined as an effort to improve teaching and learning through the targeted analysis of student data, be it performance data or demographic data (Fritz 2010; Elias 2011). The core concerns of learning analytics lie in the various investments of its possible stakeholders: better and sounder learning for students, practicality for instructors, and pedagogical and instructional efficiency of use to an institution at large (Hirst 2011). The Point of Originality ideally attempts to address the needs of each of these stakeholders by allowing instructors to quickly and accurately assess evidence of student mastery, identifying struggling students before it becomes too late.

In the sections that follow, Sect. 8.2 provides background on the problem at hand, discussing the role of iterative exercises like blogging in the education process, problems that occur with these assessments in larger gateway courses, and the prior efforts

that have been attempted to possibly correct the problem. Section 8.3 discusses a possible means of leveraging iterative assessment to evaluate the originality of student writing. This section then describes the particular function of the Point of Originality analysis method. Section 8.4 explains the design principles underlying the Point of Originality system architecture. Section 8.5, which constitutes the bulk of the chapter, then provides two case studies for the application of the Point of Originality, where each case study was applied to different educational contexts. The second of these case studies is offered as an elaboration on the findings of the first, with the Point of Originality tool applied to a larger size class, a larger number of query terms, and in a context where a more specialized and technical vocabulary was likely on display. It is this later case study that suggests the most streamlined possible use case for the Point of Originality tool and for the viability of iterative assessment and analysis generally. The chapter concludes with Sect. 8.6, which discusses the implications of this final case study, and its applications for future iterations of the tool.

8.2 Background

8.2.1 Blogging, Iterative Exercises, and Education

Blogging in a collaborative environment is an example of a social computing activity that can be very conducive to learning (Du and Wagner 2005; Larusson and Alterman 2009). Overall, blogging provides a platform that promotes individual expression, enables students to establish their own "voice" and yields a richer conversational interactivity within a community (Williams and Jacobs 2004; Wise 2005). Each student has a blog, composed of multiple blog posts. Students can read one another's blog posts and comment on them. Because blogs are easy to use, they can promote students' digital fluency (Huffaker 2005) and encourage students to explore and publish their own nascent ideas under less pressure than in the rough-and-tumble of in-class discussions (Althaus 1997).

Writing a blog forces students to become analytic and critical as they contemplate how their ideas may be perceived by others (Williams and Jacobs 2004). Being able to review older contributions affords reflection and enables students to revisit and revise their artifacts, further developing their own viewpoints in the context of each other's writing as they sense how others understand the material similarly or differently (Oravec 2002). Conversations emerge when students read, and then comment on, each other's blog posts, thus enabling them to exchange, explore, and present alternate viewpoints on the course material (Ferdig and Trammell 2004). This type of social explanatory discussion can benefit learning (Deitering and Huston 2004; Chi and Van Lehn 1991).

Alternative spaces such as asynchronous discussion forums are another example of a technology that is sometimes used to mediate online discussions between students. However, as predominantly shared community spaces, forums give students

voices that are heard but are without a distinct, individual identity (Duffy 2008). Critical thinking may emerge for individuals, but the organization does not promote the coherent and interactive dialogue necessary for conversational modes of learning in the same way that collaborative blogging does (Thomas 2002).

Exercises like blogging can serve as an opportunity for students to master course content in an iterative fashion, developing their understanding of course material before formal evaluations come due. A blogosphere can function as a repository of information, opinions, monologues, and dialogues about course content, where students participate, and leverage each other's contributions in other educational activities (e.g., when writing term papers) (Alterman and Larusson 2009, 2010). Blogging enables students to gather their thoughts and come better prepared for class (Juang 2008; Deitering and Huston 2004), and can be predictive of student performance in a course (Du and Wagner 2005). Even non-active student bloggers can benefit from the blog's educational value as it exposes them to different views of the material without necessarily participating directly (Williams and Jacobs 2004).

Overall, having students discuss and/or "argue" about course readings has significant educational utility (Reznitskaya et al. 2001; Andriessen et al. 2003; Andriessen 2006). Some discussion might take place during class; however, class time is a limited resource. This is particularly true for larger classes, or the so-called gateway courses, that typically enroll large numbers of undergraduates where there is simply not time for everyone to speak up. By blogging, students can both express individual voices and continue conversing with their peers outside the confines of the physical classroom. Unfortunately, the sheer size of these courses presents several challenges.

8.2.2 Problems with Larger Gateway Courses

The ability to monitor and respond to student progress is ever more imperative given the realities of the modern classroom. As noted even a decade ago, the political economies of American universities increasingly mandate large class sizes, particularly in the introductory or "gateway" courses that are typically a student's first exposure to collegiate work (MacGregor 2000). These large classes have a negative impact on students and instructors alike. There is, for example, an abidingly inverse correlation between class size and student achievement (Glass and Smith 1979; Smith and Glass 1980). The large lecture, while useful for reinforcing rote facts, is less successful in fostering higher-order thinking (McKeachie and Chism 1986; Cooper and Robinson 2000), or in encouraging students to construct their own understanding of core concepts (MacGregor 2000). Such a sizable student population further constrains instructors' abilities to familiarize themselves with students' individual learning styles (MacGregor 2000), thereby forcing instructors to assume that their audience consists of uniform types of learners (Cooper and Robinson 2000). Although the extent of feedback that students receive is one of the most powerful predictors of student achievement (Walberg 1984; Rosenshine and Meister

1995), instructor feedback in large lecture courses is often slow and sporadic; students typically need to wait weeks—from, for example, one midterm assessment to the next—to put their course-related skills into practice, and even longer than that to have their assignments evaluated by an instructor (Cooper and Robinson 2000). Pedagogical adjustments, in other words, become both more unwieldy and more unlikely in the precise environment where they would be most necessary.

Given the problems inherent to large lectures classes, but given also their entrenched status within the American university system, it would thus logically be prudent to find a way to minimize their most pernicious consequences. Any broader attempt to remedy the problems of larger gateway courses should thus aspire to first, foster higher-order thinking; second, to suit multiple types of learning styles; and third, to provide students with feedback as rapidly as possible. These first two objectives are inherent virtues of the collaborative blogging process; the final objective is the focus of this chapter.

8.2.3 *Prior Efforts*

Several attempts to minimize the unintended consequences of large gateway courses exist. Almost all efforts call for resizing the large class group, either by literally subdividing the class or else by designing activities to make the large class "seem" small. This latter method, it might be argued, is the one already pursued by student participation in a blogging environment, where conversations take place in an ad hoc and freeform manner.

Known interventions can be roughly classified into two major groups: those interventions that are specifically meant for in-class use, and those interventions that are intended to take place between classes. Those activities that take place during class typically interrupt the lecture itself (Mills-Jones 1999; Nicol and Boyle 2003; Brewer 2004), asking students, for instance, to respond to a series of prompts which they answer through remote devices. These same activities, however beneficial, generally disrupt the actual process of knowledge transmission and tend to reward rote memorization rather than higher-order thinking; what feedback students receive reflects only whether or not they got a prompt right or wrong, and not how well or how comprehensively they understood the material. Since the activities take place in the classroom, and in front of the entire student population, the activities themselves moreover treat all students in exactly the same manner regardless of learning style.

Interventions intended for use between classes are roundly invested in providing instructors with observable statistical modeling in near real-time (Robinson 2001; Brewer 2004; Gerdeman et al. 2007), which can then be referred to before the next session. These activities typically attempt to encourage higher-order thinking by forcing students to reflect on their own learning, asking, for instance, that students rate their level of confidence before responding to prompts (Brewer 2004), or that they engage in a collaborative peer review of one another's written work (Robinson 2001; Gerdeman et al. 2007). The benefits of this type of activity are directly

analogous to the benefits of blogging collaboratively as deployed in this chapter. In any event, what is primarily under consideration is the measurable development of higher-order thinking.

8.3 Evaluating Originality

8.3.1 Originality in Student Writing

When students engage in a writing activity, the final evaluation of their work cannot only assess whether or not the student has provided the most closely correct answer. Process is just as relevant to student writing as content (Taylor 1981). Student writing that exhibits exceptional higher-order thinking is generally seen as that which demonstrates a mastery of the course material in new, profound, or statistically unusual ways (Moore 1985). The ideal is not only for students to confirm that they've understood lectures, but to do so in ways that even the educator might not have thought of. This process of mastery need not take place all at once. As a student is continually exposed to the same material, or is given the independent opportunity to rethink, reframe, or revisit that material (Tynjälä et al. 2001), his or her writing on the subject has the chance to evolve, from rote regurgitation to wholly original expression (Nelson 2001). At the level of language, this evolution is reflected through *recasting*.

Recasting is the learning process whereby a student refines his or her understanding of a concept found in course lectures or readings by putting that concept into his or her own words (Shih 1986). In the acquisition of new languages especially, this process can be useful, because it allows students to acquire new vocabulary using the assortment of words already available to them (Shih 1986; McDonough and Mackey 2006). Even where the student's understanding of a language is not an explicit concern, recasting can mark a student's attempts to graduate to more sophisticated or professionalized terminology, or, inversely but to the same end, to place new concepts into terms that are nearer to what the student would naturally be more likely to say (Eilam 2002). It is this ability to put concepts into one's own words, discovering more "original" expression of the same concepts, that is meant by the term "originality" in the Point of Originality's name. "Originality," fully defined, can of course take numerous forms. The concept of recasting, however, spans a number of theoretical orientations, with an influence on theories of schema formulation (Korthagen and Lagerwerf 1995), the sensemaking process known as "scaffolding" (Gee and Green 1998), as well as the express principles of educational constructivism (Lebrun 1999).

For an instructor, the simple identification of recast terminology within a student's written work can provide an effective barometer for pedagogical self-reflection. If a subset of terms or concepts is deemed vital to the syllabus, repetitions and recast iterations of those same terms will at least suggest that those terms are being acknowledged and reflected upon. Although the presence of recast terminology

is not the only metric representative of a student's mastery, the central role that recasting plays in a host of pedagogies (e.g., Korthagen and Lagerwerf 1995; Gee and Green 1998; Lebrun 1999) suggests that writing demonstrating high or low levels of recasting will reflect other aspects of performance within the course. Yet if the instructor hopes not only to identify instances where key concepts are deployed, but to determine how comprehensively the concepts are being internalized, it is first necessary to possess a method of scoring how original any given recast might be. In order to do this, we have developed a metric for isolating a specific *point of originality* within student writing.

8.3.2 The Point of Originality Analysis Method

The process of evaluating student writing in terms of wholly original expression is primarily composed of two parts: the analysis method and a custom-made visualization depicting each student's "originality" at any given time throughout the duration of the semester.

8.3.2.1 Analysis Method: Theoretical Background

To identify whether students are deploying recast terminology related to an individual course concept, it is first necessary to determine the full range of ways in which other words might relate to that concept. The Point of Originality method uses WordNet, a lexical database that arranges nouns, verbs, adjectives, and adverbs by their conceptual-semantic and lexical relationships, for just this purpose (Fellbaum 1998). Whereas a simple thesaurus would be able to identify any two words as synonyms or antonyms of one another, WordNet is able to note the similarity between two words that don't have literally identical meanings. These relationships are ideally meant to mirror the same lexical associations made by human cognition.

WordNet's arrangement is hierarchical, which is to say that certain terms are more closely related than others. Within WordNet, these relationships are displayed as "synsets," clusters of terms that fork, like neurons or tree branches, from more specific to more and more diffuse associations (see Fig. 8.1). If two words are found within one another's synset tree, it stands to reason that these terms are, in some way, related, be it closely or distantly. As discussed in the next subsection, the distances between two terms can be calculated, and assigned a value commensurate with their degree of semantic relatedness (Boyd-Graber et al. 2007).

The hierarchical arrangement inherent to WordNet provides one method of determining the relationship between two terms. If the synset tree of one term encompasses another term, it is simple enough to note how many synset jumps it takes to move from one to another. In Fig. 8.1, a "Dalmatian" is a type of "dog," which itself belongs to the subcategory of "domestic animals;" thus there are two

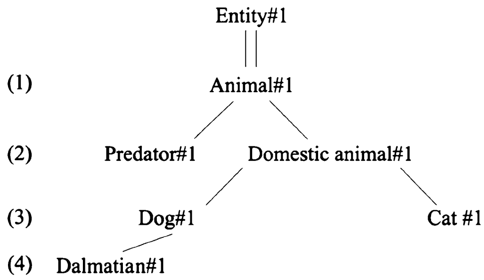

Fig. 8.1 Model synset tree (by hyponym relation)

Table 8.1 Possible lexical entailments for nouns in WordNet

Synonym: X is a synonym of Y if X *means* Y
Example: {smile, grin}
Hypernym: X is a hypernym of Y if every X is a *kind of* Y
Example: {dog, mammal}
Hyponym: X is a hyponym of Y if every Y is a *kind of* X
Example: {mammal, dog}
Holonym: X is a holonym of Y if Y is *part of* X
Example: {hand, finger}
Meronym: X is a meronym of Y if X is *part of* Y
Example: {finger, hand}

tiers of associations between the concepts of "Dalmation" and "domestic animals." Unfortunately, however, just how closely any two terms might be related is not a purely linear relationship. WordNet organizes related terms by their precise lexical entailment, such that nouns might be categorized as synonyms, hypernyms, hyponyms, holonyms, and meronyms, as seen in Table 8.1.

These possible entailments provide a rudimentary roadmap for all the ways in which two words might be related. Since WordNet attempts to map the cognitive associations automatically formed between words (Fellbaum 1998), a student's evocation (Nikolova et al. 2009) of the holonym or hypernym of a given noun instead of the noun itself is more likely to form an associative recast of the original term. If a core concern of a course was with the term "democracy," for instance, a student's ability to discuss the antithetical concept of "tyranny" would be an indication of the student's nuanced appreciation of the original term.

Yet while this simple index displays just how any two terms might be related, all the possible relationships noted are not necessarily equal. Some relationships, like that between synonyms "smile" and "grin," are obviously bound to be more strongly associated than that between "mammal" and "dog." Following a method first noted by Yang and Powers (2005), it is possible to install a series of weights that can best calculate the semantic distance between any two terms. This method in particular is useful because of all known methods, it bears the highest correspondence between its own distance calculations and the intuitions of actual human respondents (at 92.1 % accuracy).

8.3.2.2 Analysis Method: Implementation

Determining the *point of originality* of a student's blog post depends upon the manual input of a specific query term by the instructor. A query term corresponds to a specific key course topic. Having the instructor manually input the query term reinforces the pedagogical utility of the process, for it's the instructor, foremost, who will finally be responsible for selecting the most relevant topics of the course. For a specific query term, the Point of Originality method generates a WordNet synset tree. Words within the tree are then compared to the body of words extracted from a student's blog post. Where matches are found, a summation of distance calculations between the original query term and the matches is performed as follows:

Let q be a query term supplied by the instructor. Then, let $W = \{w_0, w_1, \ldots, w_n\}$ be a set containing all synset word matches (w) from the WordNet database for q.

Let $B = \{b_0, b_1, \ldots, b_n\}$ be a set of all words composing a blog post by a particular student and let $S = \{s_0, s_1, \ldots, s_n\}$ be a set of stopwords, a list of common words in English usage (like "the" or "and"), to be omitted to speed up processing time. Then, $M = \{m_0, m_1, \ldots, m_n\}$, the set of synset term matches found in a blog post for query term q can be defined as:

$$M = W \cap (B - S)$$

WordNet stores synset matches in a tree structure with q as the root node. Then, δ, the distance (depth) for any given synset match ($m \in M$) from the root node (query term q) is defined as:

$$\delta = \begin{cases} 0, & \text{if } m = q \\ 1, & \text{if } m \text{ is first child of } q \\ 2, & \text{if } m \text{ is second child of } q \\ 3, & \text{if } m \text{ is third child of } q \\ 4, & \text{if } m \text{ is fourth child of } q \\ 5, & \text{if } m \text{ is fifth child of } q \\ 6, & \text{if } m \text{ is sixth child of } q \end{cases}$$

WordNet also supplies the lexical entailment of each synset term. Thus, t, the "word type" of any given synset term match $m \in M$, is defined as:

$$t = \begin{cases} 1.0, & \text{if } m = q \\ 0.9, & \text{if } m = \text{synonym / antonym} \\ 0.85, & \text{if } m = \text{hypernym / hyponym} \\ 0.85, & \text{if } m = \text{holonym / meronym} \end{cases}$$

Then α, the weight of any given synset term match is calculated as:

$$\alpha = (\delta \times 0.7) \times t$$

The depth for any given synset term is multiplied by a constant value of 0.7, which reflects the diminished associations between two terms the farther separated they are along the synset tree. This value is selected because it corresponds with the calculation of distance between terms that yields the nearest match with human intuition (Yang and Powers 2005).

Then, C, the cumulative originality score for a given query term q in a student's blog post, can be defined as:

$$C(q) = \sum_{n=0}^{|M|} \alpha_n$$

The *point of originality* for a particular course topic is in many cases defined by the presence of several related query terms, or in other words, the synset matches for those terms. By defining $Q = \{q_0, q_1, \ldots, q_n\}$ as the set of query terms supplied by the instructor at any one time, then P, the overall *point of originality* of a given student's blog post for a particular course topic (defined by Q), is:

$$P(Q) = \sum_{n=0}^{|Q|} C(q_n)$$

Finally, repeating the above *point of originality* calculation for each blog post written by a particular student, and plotting all instances of originality on a horizontal timeline, allows for an optimal instruction comprehension so that the instructor can see recasts of a particular course topic (defined by Q) across the entire body of a student's writing throughout a single course.

Although this chapter focuses on the analysis of blog posts as students' writing samples, given some additional programming work, any electronic form of student writing could be made compatible with the tool for subsequent analysis provided it could be captured in a chronologically ordered RSS feed, as discussed in greater detail in Sect. 8.4.

8.3.2.3 Visualization for the *Point of Originality*

The timeline visualization, as seen in Fig. 8.2, displays a horizontal timeline that represents the time interval for all the writings of any student for the duration of a particular semester. The numbered components of Fig. 8.2 correspond to the following features.

1. This drop-down menu allows the instructor to select which student's writing samples are currently being displayed.

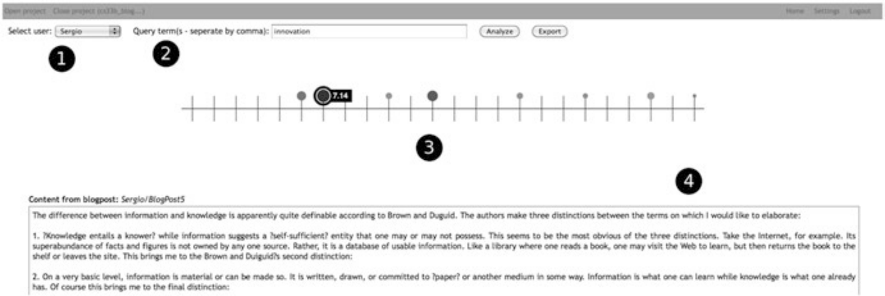

Fig. 8.2 The point of originality timeline visualization

2. This is where query terms (Q) are input by the instructor.
3. This timeline displays the date/times of each of the student's writing samples. Each marker is color-coded, from colder to warmer colors along the ROYGBIV spectrum, the higher the value of the *point of originality* (P) score for any given writing sample. These color assignments present an intuitive way for the instructor to quickly recognize that the sample has been assigned a higher originality value.
4. If a writing sample marker is selected in the timeline window (see inset 3), the text of that writing sample is displayed here.

This assortment of visualization options allows the *point of originality* calculation to be displayed in a number of intuitive ways: both within chronology (inset 3) and in context (inset 4).

8.4 System Design and Architecture

The Point of Originality tool consists of three primary parts, two of which, the analysis algorithm and custom visualization, were described in preceding sections. The final and capstone part of the system is a web application that couples the analysis algorithm and visualization platform with a basic web-based interface, allowing a user to initiate analysis, interact with the data, and interpret the results.

8.4.1 Overview of System Architecture

The entire Point of Originality tool, from back-end algorithmic analysis to front-end web interface and visualization, is written in the Python programming language ("Python"; Van Rossum and Drake 2003). Python was chosen, foremost, because it is highly applicable to scientific computing (Rashed and Ahsan 2012). Python itself is an open source, object-oriented, high-level language that is relatively easy to

learn and with a large, active user base, making passing the "development baton" somewhat easier than with other languages. Moreover, with a largely human readable coding syntax, significant improvements are gained during the prototyping phase, with less time required to focus on the internals of coding. This is especially useful since user time is more valuable than computational processing time during preliminary phases of scientific computing (Rashed and Ahsan 2012).

Early in the project, it was decided to build the tool as a web-based application in the cloud. Building the Point of Originality tool as a web-based application, which could be easily accessed via a standard web browser, would alleviate most resource constraints, system maintenance overhead, and upkeep from the kinds of users envisioned for the system—instructors—so that attention could be focused solely on analysis and intervention.

As a web application independent of any specific operating system, adoption of the tool in any number of course, instructor, and institutional settings is dramatically simplified, as the only requirements for implementation is an Internet connection and a standard web browser. To optimize compatibility with a variety of different teaching and learning systems, writings can be uploaded to the Point of Originality tool in a simple RSS format that organizes student writings in a standard, chronological, and uniform manner.

In short, using Python as the primary programming language for the entire project enabled the building of the analysis algorithm, the data visualization, user interface, and interaction modules all in a single homogeneous infrastructure.

The architecture design of the system follows the common Model-View-Controller model, separating those parts of the systems that handle the data and business logic, the mediation of input, and user interaction and representation of the data from one another (Krasner and Pope 1988).

In the standard fashion, the user manipulates the UI controls to select and filter the dataset to be analyzed and inputs the appropriate query terms. These preferences are then processed by the controller, which retrieves, filters, and manipulates the data stored in the model, which then updates the visualization of the original dataset in turn. The cycle repeats itself as needed. This design pattern was chosen as it applies well to both the system architecture used, programming language deployed and the type of data being analyzed.

8.4.2 The Analysis Algorithm

To improve performance, scalability, and stability of the analysis algorithm and overall system, some data preprocessing tasks that the algorithm would require takes place when the data is first uploaded. This section will briefly emphasize the utility of those tasks.

When a set of writings is uploaded, the system begins by parsing each individual writing sample, creating an entry for that writing in the database. The database used for the project is SQLite ("SQLite"), that although an ordinary disk file provides

many of the more advanced features typically found in larger database servers when optimized.

As the system parses the writing samples, in addition to storing an original copy of the writing text in the database, it also creates an abbreviated copy of each writing text by removing all stopwords, leaving only those words that are of immediate relevance to the instructor (and, consequently, the originality algorithm). Therefore, the unit of analysis for the algorithm is actually a smaller subset of words than the original writing sample. Furthermore, during analysis, the algorithm does not need to locate each occurrence of a query term in a writing. What the algorithm needs to know is only (a) did a query term appear in a given writing and (b) how often. Therefore, after removing stopwords, the database engine deploys a common technique where an index or list of search terms of the textual content is built. Only this index is used during analysis, providing an extremely fast way of retrieving what words appeared in a text, where, and how often.

These two preprocessing tasks drastically reduce the runtime of a single originality analysis as they remove the need for the algorithm to cycle through and compare every query term to every word in every writing by every student. Having the ability to recall this information quickly and efficiently is important because of the possible complexity of the relationship established between query terms and query term hits during analysis. Once an instructor has generated originality scores for a specific set of query terms, the results cannot simply be "recycled" if the instructor wants to add or subtract even one additional query term, because the weights of the individual terms would have to be reassigned as the lexical dependencies change. In the event that an instructor does need to rerun a previously used set of query terms for analysis, the system will retrieve the full results from the database rather than computing their values again.

8.4.3 Other Technologies

The web application was built using Flask ("Flask"), a lightweight Python-based web framework that provides a collection of libraries that promote code reuse and alleviate the overhead associated with common web development activities, speeding up both the development and prototyping phases. A number of other frameworks were used as well to build various parts of the application. These are summarized in Table 8.2.

8.5 Case Studies, Results, and Analysis

Two separate studies of the Point of Originality system have been conducted at present, the first of which (Larusson and White 2012) was intended as a proof of concept to test the accuracy and efficacy of the system. Since the analysis generated through

Table 8.2 Overview of toolkits and frameworks used

Framework	Purpose
SQLite	Database engine used to store student writings
WordNet	Lexical database used to construct the synset trees that are used to calculate originality
Natural Language Toolkit	Used to access WordNet and work with and manipulate textual data through classification, tokenization, stemming, tagging, and parsing
Flask	Web framework used to build the web application through which the originality analysis is conducted
Graphael	JavaScript library used to build the originality timeline visualization
jQuery	Used to handle client side scripting of HTML and asynchronous data retrieval
Bootstrap	Front-end framework used for typography and designing forms, buttons, tables, grids, navigation, organization, and everything else related to presentation of web content
SQLAlchemy	An open source SQL toolkit and object-relational mapper (ORM) for the Python programming language for data manipulation

the Point of Originality tool is meant to be predictive of student success or failure, it was imperative that the Originality algorithm could be readily applied to an existing dataset to determine whether the results obtained by the system could identify trends in student behavior before formal grading had taken place.

The second study sought to broaden the scope of inquiry that the Point of Originality system was capable of conducting. An exercise like the one conducted in the original study would require an instructor to generate a rough list of query terms in advance, and would use those query terms only to assess likely student achievement for a corresponding summative exercise that had been concerned with those topics. While this use of the system was undoubtedly of interest, it remained to be seen whether an instructor could use the Point of Originality tool more holistically to predict student success in a course irrespective of any single exercise and without specification of individual query terms. For this case study, rather than using a predefined list of query terms that would be associated with a later summative exercise, we've chosen to perform what we could instead think of as a "mass analysis," using 112 distinct words culled from a course syllabus. The means of analysis remain identical, yet the results obtained are intended to provide a survey of student originality for every possible term that would be central to the course.

This second case study also examines a very different kind of course from the prior study, in an attempt to expand the domain of use cases to which the Point of Originality system might apply. Whereas the earlier study had focused on a computer science class containing only 25 students, and which explored such topics as intellectual property, virtual communities, and privacy, this study will focus on a class in the hard sciences, specifically neuroscience and neuropsychology, with a class population of 71 students. Since one of the primary intended uses of the Point of Originality tool is to provide oversight of large gateway courses where the output of students is assuredly beyond the immediate oversight of any single instructor, the

authors needed to determine whether the system's insights could scale to an expanded course size. Selecting a course in the hard sciences, however, presented an additional challenge and an additional opportunity, for it would force the tool to be applied to an environment ostensibly less liable to require discursive discussion. Although students had engaged in a blogging activity similar in design to that of the earlier study, the range of topics under consideration could be expected to contain a more specific and technical vocabulary than the system had yet attempted to analyze. Rather than relying on topically relevant but still fairly broad query terms like "innovate commons" or "layers, resources, and control," the kinds of query terms required here would involve scientific terminology such as "cerebellum," "limbic system," and "neurotransmitters" alongside discussion of medical conditions like "autism" and "Parkinson's disease." The possible challenge of such a dataset would lie in the nominally limited lexical variability of such terms. There simply aren't many good substitutes for a word like "neurotransmitter." The accuracy of any results obtained by the Point of Originality system would thus be tested by the constrained conditions of the course itself.

Our eventual results not only reaffirmed our belief in the efficacy of the Point of Originality system, but actually suggested *superior* performance when applied to such a course context in this way.

8.5.1 Phase I: From Iterative to Evaluative—Originality in the Lead-In Period

This section reports on a case study that explores the capability of using the Point of Originality tool to assess the originality of student writing in a semester-long blogging activity. More specifically, the study focuses on correlating originality scores assigned to students' blog posts with both their activities in the blogosphere during the semester and the final grades assigned to a term paper covering the same topics. Although primarily aimed at testing the validity of the Point of Originality method, this study models a likely use case. By demonstrating how low *point of originality* values correspond to poor performance in other aspects of the course, the Point of Originality tool could provide instructors with an early, near-instantaneous diagnostic of which students might require additional help. The tool might thus ideally streamline the process of conducting targeted pedagogical adjustments or interventions.

The blogging data was collected from a course titled "Internet & Society." This course was taught in the Fall of 2008 in the Department of Computer Science at Brandeis University. The course is an *introductory course*, focused on exposing students to topics such as the social life of information, virtual communities, privacy, intellectual property, and peer-to-peer computing.

In the collaborative blogging activity, each student has a blog where he or she writes opinions on the course readings. Students can read each other's posts and comment on the posts of their peers. The blogosphere provides several features focused on increasing students' awareness of recent activity, and enabling them to find interesting blog posts to read and conversations in which to participate.

8.5.1.1 Participants

There were 17 male and 8 female students enrolled in the class, all of whom were undergraduates. There were 3 science majors and 1 science minor in the class, with 12 students majoring in the social sciences and 8 minoring in the social sciences. The remainder of the class was either in the humanities or fine arts. Three students were omitted from the dataset because they did not begin blogging until the final weeks of the course following a warning from the instructor.

As an introductory course, readily open to non-majors, there were few technical requirements for enrolling. No formal evaluations were conducted to assess students' computer literacy or prior knowledge of the domain. In class discussions, most students expressed moderate to advanced technical skills.

The instructor and teaching assistant did not design or implement the blogging activity in such a way that it could potentially influence the students' choice of topic or writing styles. No students were assigned to preexisting authoring roles, and all students were simply expected to conform to the standards of the assignment as follows.

8.5.1.2 Procedure

At the beginning of the semester, an in-class exercise introduced students to the important features of the blogging environment. There were two course lectures per week, and students were required to create at least one blog post per lecture. Posts were typically 1 or 2 paragraphs long. Students were also required to read and comment on other contributions to the blogosphere. The blogging work of each student counted for 35 % of the final grade.

During the semester, the students read four books and wrote a paper on one of these books. The focus of the analysis presented in this chapter is on the blogging work that the students performed while the class was still reading and discussing concepts from the book about which they would eventually write papers.

8.5.1.3 Metrics

Midway through the semester, students were asked to write a graded essay on the topic of the "innovate commons," and how this concept related to other course discussions of "layers, resources, and control." To determine how well students had mastered these individual concepts prior to writing their graded essays, these five words became the query terms used for analysis.

8.5.1.4 Method

All blogging activity from each of the students was automatically recorded in a transcript and analyzed using the Point of Originality tool. Originality scores were generated by the tool for all blog posts as well as all papers. These results were then

correlated, first, with students' final paper grades, and separately with statistical data suggesting the way in which students had used the blogging environment during the blogging part of the semester.

8.5.1.5 Results

The analysis was composed of two principle parts.

The first part compared the degree to which the tool indicated the originality of the students' blog posts and how well the originality scores related to the grades that the instructor assigned to their papers. In the ideal situation, given that the instructor graded their papers based on how well the students expressed higher-order understanding of the course material, (or, in other words, whether their writing reflected original thought), the tool should provide scores where higher originality values would correspond to higher paper grades.

The second part sought to explore to what degree the students' interactivity in the blogosphere influenced their understanding of the course readings, and in what way their immersion in the blogging community positively or negatively impacted their levels of originality when writing papers. Ideally, students would find sufficient impetus to become deeply involved in the blogging learning community, and their exposure to alternate or similar viewpoints of the same materials would help them to develop their own viewpoints or to strengthen existing ones, thus leading to more original thought and better papers.

Since the analysis was primarily concerned with ensuring that the tool could be used during a course to preemptively diagnose likely student success, the blog post dataset was filtered to only include blog posts written in what was defined as the "lead-in" period of blogging. During this period, the students were writing blog posts and comments on the topics that they eventually wrote their papers on, but at the time were unaware *which* specific topics they would have to address in those papers. The paper grades were assigned during the fall of 2008, roughly two years prior to the study described in this chapter. Furthermore, grading was done by the course instructor, who is not a direct participant in the Point of Originality project.

Originality in the Lead-In Period

We began by collecting the originality scores calculated by our system for the blog posts written by each student on the paper topics along with the actual grades that each student received for his or her paper. The average grade for student papers was 80.00 with a standard deviation of 16.83. The highest grade assigned was 95 and lowest was 40 on a scale from 0 to 100.

The students' blog posts received on average an originality score of 10.61 with a standard deviation of 4.29. The highest originality score assigned by our system was 18.30 whereas the lowest score was 3.92.

Table 8.3 Originality variance and paper grades for two different groups of students

Metric	Average	SD	SEM	*N*
Above average grade				
Paper grades	90.63	3.20	1.13	8
Originality variance	−6.10	21.92	7.75	8
Below average grade				
Paper grades	66.79	15.14	4.05	14
Originality variance	21.49	27.63	7.38	14

Soon, a pattern emerged indicating that the more original the students' blogging work, the higher the paper grades assigned by the instructor. This first finding proved pivotal in determining that the Point of Originality tool was automatically producing results that would potentially correlate to standard approaches to pedagogy. A Pearson correlation coefficient test confirmed that there was indeed a positive correlation between the two factors. As students' blog post originality scores increased, their final paper grades covering the same topics increased as well. In other words, as their blogging activity became more original, the students wrote better papers:

$$r(20) = 0.492, p = 0.05$$

To further confirm the potential relationship between originality while initially learning the course materials (during the lead-in period) and how well that work transformed into mastery of course content as reflected by paper writing, students were divided into two groups based on their paper grades. Students whose paper received a grade above the average (80.00) were assigned to one group, the *upper* group, whereas students who scored below the average were assigned to the *lower* group.

As shown in Table 8.3, the students in the *upper* group received an average grade of 90.63 on their papers whereas the students in the *lower* group received an average grade of 66.79. What is more interesting, however, is what can be defined as the *originality variance*: the difference between how original the students' blog posts were compared to their final papers. While the *lower* student group had an originality variance of 21.49, the variance for the students in the *upper* group was −6.10.

Because the variance for the *upper* group is negative, those students' blog posts, written during the lead-in period, were on average more original than their final papers. Although it might seem then that those students were not necessarily more original than the students in the *lower* group that is not, however, the case. The fact that the variance is negative for the *upper* group is indicative of the fact that those students were at the *height* of their understanding of the materials *even during* the lead-in period. These students had mastered the materials in such a way that they had an easier time writing their papers, whereas the students in the *lower* group were only first beginning to wrestle with this content after the papers were assigned. This is suggested by the fact that the originality variance for the *lower* group was a positive value of 21.49, a value more than twice as great as the students' average originality score during the entire period.

A t-test of independent samples confirmed that the originality variance between the *upper* and *lower* groups was indeed statistically significant. Students who had received higher grades for papers wrote blog posts that were more original in the lead-in period:

$$t(20) = 2.42, p < 0.02$$

The key observation is whether or not students' retention of course materials was equal for both groups. Students that master materials for the first time only when preparing for graded assessments don't necessarily have the ability to apply that knowledge after the course ends because their "grasping" of the content was short lived. These students had never exercised their understanding of the material prior to being evaluated. If students can get "into the game" earlier in the semester, they have greater opportunities to participate in discussions, refine their understanding and "lock it down deep" so that they leave the course with a higher degree of mastery.

In a large reading- and writing-intensive course, where a bulk of the work towards mastery might take place in machine-readable form, it goes without saying that it would be advantageous for the instructor to be able to use technology to monitor each student's progress. Specifically in larger gateway courses, where the odds are already stacked against student achievement and the need for interventions is more difficult to spot, students who fail to integrate completely with the class community—either because their experience comes from another discipline, or because they simply aren't accustomed to the specific class environment—are likely to suffer poor performance. Having the ability to assess students' mastery of the material, however, would enable the instructor to identify those students who are perhaps struggling or only falling behind, and to intervene to correct the students' performance.

Interactivity in the Blogosphere

In an online technology-mediated community like the one described in this case study, students benefit from exposure to both similar and contrasting viewpoints of the same course material. If the students' deep emersion in the blogging activity has a positive impact on their learning, one can assume that the originality score would correlate with the degree to which each student participates online. In other words, higher originality scores should correlate with positive student outcomes for those students that take advantage of the technology-mediated activity, frequently reading other students' viewpoints and partaking in thoughtful conversations about the course readings.

To assess student participation in the blogosphere, each student's *exposure* (reading blog posts and comments by others) and *contributions* (writing blog posts and comments oneself) were measured. These activities were then correlated with the originality scores assigned to each student's paper. Table 8.4 summarizes these metrics.

Table 8.4 Originality and interactivity in the blogosphere

Metric	Average	SD	SEM	*N*
Paper originality score	53.49	14.53	3.10	22
Exposure	4.36	3.93	0.84	22
Contributions	4.18	2.17	0.75	22

Overall, the student papers received an average originality score of 53.49, with a standard deviation of 14.53. The highest originality score was 93.76, whereas the lowest score was 31.66.

In terms of exposure in the blogosphere, the average number of times that a student was exposed to other students' contributions was 4.36, with a standard deviation of 3.93. The highest number of contributions read by a student in the blogosphere was 14, whereas one student read no contributions by the class at all. A Pearson coefficient correlation was used to explore the potential correlation between the originality of student papers and the degree of each student's exposure in the blogosphere. As shown below, there is a statistically significant positive correlation between the two factors. In other words, higher exposure in the blogosphere led to more original papers:

$$r(20) = 0.44, p < 0.05$$

In terms of contributing in the blogosphere, each student made on average 4.18 contributions during the lead-in period, with a standard deviation of 2.17. The highest number of blog posts and comments written by a student was 9, whereas the lowest number of contributions was 1. As before, a Pearson correlation test confirmed that there was a statistically significant positive correlation between the number of contributions a student makes in the blogosphere and the eventual originality of his or her paper.

$$r(20) = 0.42, p < 0.05$$

8.5.2 Phase II: Amplified Analysis—Applying the Point of Originality to a Large Class in the Hard Sciences

8.5.2.1 Participants

The data was collected from a Neuropsychology (NPSY) class taught at Brandeis University during the fall of 2009. There were a total of 71 students enrolled in the class.

The instructor and teaching assistants did not assign students to any predetermined authoring roles. Occasionally the instructor and teaching assistants would be active in the blogging environment as well, either by commenting on

students' postings or uploading relevant non-curricular material from time to time. So that only student content is under consideration, these parties have been removed from the data set for the purposes of this study.

8.5.2.2 Procedure

Students were asked to create at least two original blog postings and one comment on another student's post per week. Blog posts were intended to run about two paragraphs in length, and students were asked to "summarize in [their] own words the key content or idea(s) of the week's reading, or develop an argument on an issue that was discussed during a class meeting." (Quotations like this one are extracted from the course syllabus or other teaching materials.) Students were told that their activity in the blogging environment would account for roughly 15 % of their final course grade.

8.5.2.3 Metrics

To create a list of query terms that would apply to every key concept that the course had covered, we systematically went through the course syllabus, identifying every concept that occurred in the course description as well as those that occurred as separate items of discussion from week to week. A preliminary description of the course stating that "[t]he field and this course focus on neurons, brain structures, and neural function that are the biological foundation of the mind" thus produced the query terms "neuron," "brain structure," "neural function," "biological foundation," and "mind." Every nongeneric noun was culled from this course description and from the course syllabus. This ultimately resulted in 112 unique query terms to be used as inputs in the Point of Originality system, which can be found in Table 8.5 below.

Table 8.5 Query terms used for analysis of neuropsychology course

Phase II query terms
Biology, mental, brain, mind, neuron, brain structure, neural function, idea, perception, memory, action, decision, interaction, thought, ensemble, temporal scales, spatial scales, ion channels, spikes, transmitters, synapses, nervous system, nerves, organization, sensation, object recognition, learning, control, attention, autism, disorder, neuroscience, neurotransmitter, cerebral cortex, hippocampus, memory, temporal lobe, neuroimaging, occipital lobe, frontal lobe, MRI, fMRI, learning, amygdala, prefrontal cortex, synapse, concussion, action potential, parietal lobe, thalamus, vision, synaptic transmission, basal ganglia, glia, EEG, mistakes, Parkinson's, studying, contralateralization, limbic system, audition, strategies, Alzheimer's, DTI, cerebellum, blood–brain barrier, somatotypy, ion channel, knockout, hypothalamus, nerve gas, node of Ranvier, neuron doctrine, postsynaptic, presynaptic, retina, synaptic cleft, gyrus, cerebrum, summation (temporal/spatial), declarative, neurite, gene therapy, Sylvian fissure, medulla, sulcus

8.5.2.4 Method

Students' blogging activity was automatically input into the tool through the blogging site's RSS feed, and subsequently analyzed using the Point of Originality tool. Originality scores for all of the 112 query terms were calculated for each post by each student. This produced a total originality value intended to reflect each student's original engagement with every aspect of the course material on a post-by-post basis.

8.5.2.5 Results and Analysis

Originality and Achievement

Whereas earlier work on the Point of Originality system with a smaller subset of query terms, in Phase I, had required us to define a specific "lead-in period," which would cover only the range during which students would have been exposed to those particular concepts, the more comprehensive query term list allowed us to remain agnostic as to what aspect of the course material would be emphasized from one moment to the next. Although the first weeks of the course focused on the biological components of the nervous system, where later weeks focused on specific medical disorders, our distribution of query terms ensured that all aspects of the material were continually being evaluated from week to week. Trends in originality values could thus serve as a more proximate measurement of a student's mastery of the course material most broadly.

As a means of understanding students' ability to apply course concepts over the full duration of the course, the metric of most interest was the average originality exhibited by each student over the course period. Students' average originality fell within a relatively small window. For each of the 112 query terms used as inputs, students produced an average originality value ranging from 69.63 to 182.39. The average value for average originality (the "average of the averages," so to speak) was 128.84.

These average originality values were then compared with students' final grades in the course, as given on a conventional 100 point scale. A Pearson correlation test confirmed a highly significant positive correlation between average originality and student grades:

$$r(69) = 0.336, p = 0.004$$

Put differently, the results obtained suggested that students who were more original as determined by the Point of Originality system on average, were more likely to succeed in the course, with a less than 0.4 % chance that the distribution was simply the product of random occurrence.

Although average originality over the entire course period was our most important metric for replicating an ideal use case for the Point of Originality tool, we also

wanted to account for how "concentrated" a single student's originality might be in even a single post. To this end, we wanted to look at the highest originality values that students produced over the span of the course. The highest originality value produced by each student over the span of the course covered a larger range than for the "average of averages"—running from 103.52 to 478.96, with an average high value of 249.90. A separate Pearson correlation test confirmed that exceptionally high individual originality values also positively correlated with students' grades:

$$r(69) = 0.270, p = 0.0023$$

For the sake of comparison, however, we wanted to determine just how much variance might be found between a student's average and high originality values. While high originality values, on average, were most closely reflective of student success, it was possible that freak occurrences—sudden swings in a single post's originality value, when a student had otherwise produced only middling originality—might have interfered with the data. In directly comparing average originality to highest originality value, however, a Pearson correlation test identified a near perfect correlation at the 0.0 % level:

$$r(69) = 0.745, p = 0.000$$

What this finding suggests, essentially, is that the chances of the system producing outliers are fairly low. The students who are more original on average also tend to be the students who produce the single moments of the most potent originality.

The consequences of this analysis for a likely use case are immense. At any point as a course proceeds, an instructor could productively consult the originality values for all the students in a class to identify which students remained above and which students had fallen below average. An example of what these data points look like independent of the Point of Originality's ordinary visualization method can be seen in Figs. 8.3 and 8.4. The dotted black line running across both graphs shows the average originality value for all 71 students up until that particular point in the course. (This line has a slight upward trajectory before leveling off because rather than representing a "hard" average, the line is instead simply the average originality of those posts that would have been produced through any given interval; this approximation of average originality thus more closely approximates the average originality values that an instructor using the Point of Originality tool would encounter from week to week.)

The students in Fig. 8.3 are the nine students who received the highest final grade in the course, and who were at least one standard deviation above the average final grade. The students in Fig. 8.4, meanwhile, are the seven students whose performance placed them at least one standard deviation below the average final grade. What is clear is that those students in Fig. 8.3 produce a majority of posts well above the rest of the class's average originality, whereas those students in Fig. 8.4 almost never climb above the average originality of their peers. (The single most

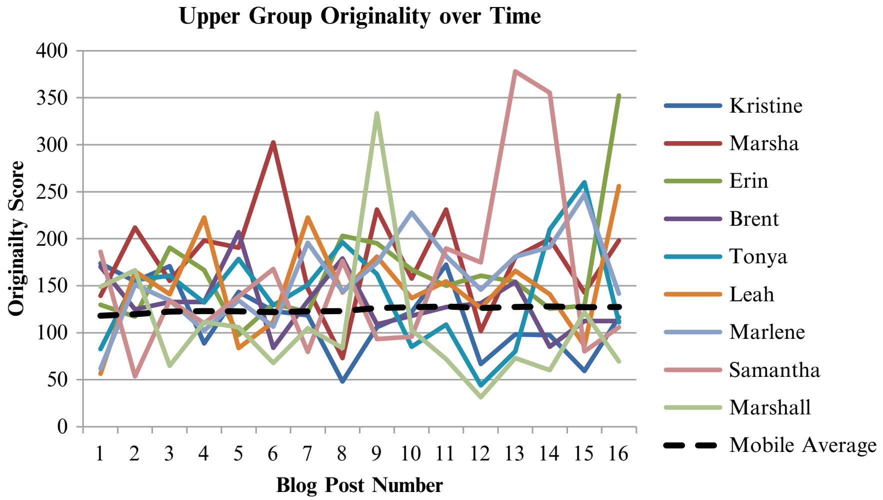

Fig. 8.3 Upper group originality over time

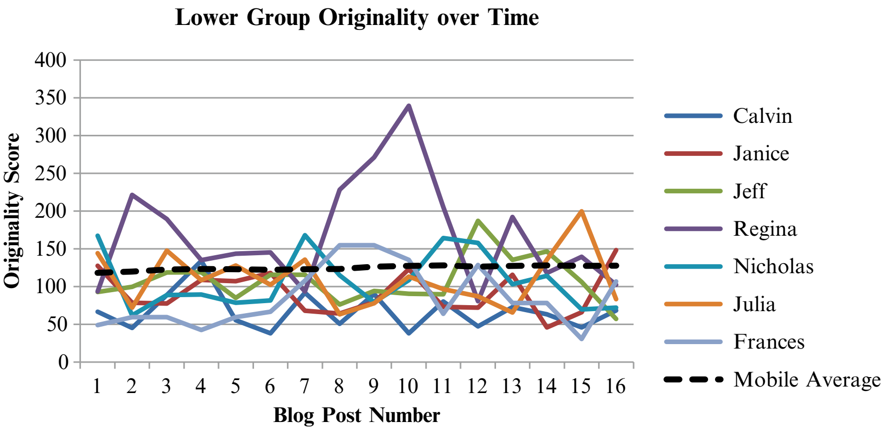

Fig. 8.4 Lower group originality over time

pronounced outlier, "Regina," is addressed more closely in the next section.) In a practical use case, the difference between these groups of students would thus be immediately available to an instructor from a course's first weeks.

Continuing Originality

The findings above would suggest that the correlations between originality values and eventual student success can be found at almost any point in a course. If a student, as for any of the students in Fig. 8.4, fails to consistently create posts more

original than the average originality of his or her classmates, it is likely that some kind of specific intervention would prove necessary. Merely by keeping an eye on students' average originality, an instructor can make these determinations quickly and on the fly, with no further need to produce new query terms from week to week.

However, we wanted to use the apparent accuracy of our results to determine whether any other trends predictive of success could be detected. Good students are those students who tend to produce posts more original than their peers on average, but we wanted to know whether the performance of those students changed significantly over time. It might be suggested, for instance, that those students who eventually did well in the course (and, consequently, had produced consistently high originality values) were actually those students who had come into the course with an already firm grasp on the material, and were thus able to produce a string of highly original posts from the beginning of the course before settling back into a comfortable level of lower effort. Or, conversely, it could be suggested that the best students were those students who peaked late, creating their most original posts as more and more of the course material had been covered in class.

To test for either of these hypotheses, we ran a regression test on each student's originality scores to determine the slope of his or her originality values over three different periods: the first half of the student's posts, the second half of the student's posts, and the full total number of posts. This method of looking at the data had the interesting consequence of temporarily rendering our earlier observations invisible; because slope deals with change over time rather than magnitude, the actual value of students' originality scores, whether high, low, or average, became functionally irrelevant.

What we found was that the slope of a student's originality in the *second half* of the class correlated positively with that student's final grade, as confirmed by a Pearson correlation test:

$$r\left(69\right) = 0.273, p = 0.021$$

Put in different terms, students who do well in the course are not only, as per our last set of results, those students who are more original than their peers on average, but those students who are able to sustain and even accelerate that level of engagement even into the final weeks of the course. Considering that our set of query terms already included all the concepts that the course would conceivably cover, such accelerating improvement is actually quite a feat, because it means that highly original students would actually need to keep topping their original success. Rather than simply demonstrating an early mastery of the material that the course will never need to improve upon or challenge, successful students only become increasingly original as the course goes on.

A representation of these findings can be found in Figs. 8.5 and 8.6. Figure 8.5 displays the originality values and originality slope of a single student, "Samantha," who earned one of the highest grades in the class, over three intervals throughout the semester. Figure 8.6 displays in turn the originality values and slope over the same three intervals for a student, "Regina," who was among the lowest performing

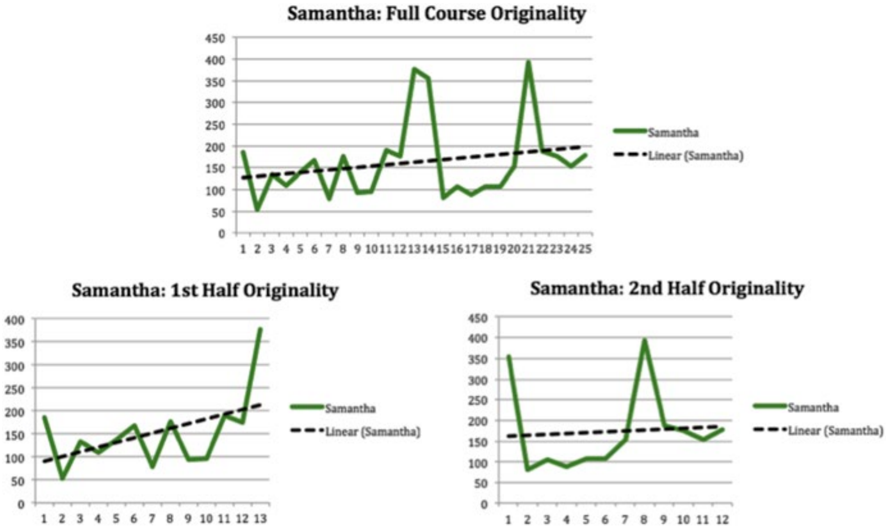

Fig. 8.5 Three originality slopes for Samantha

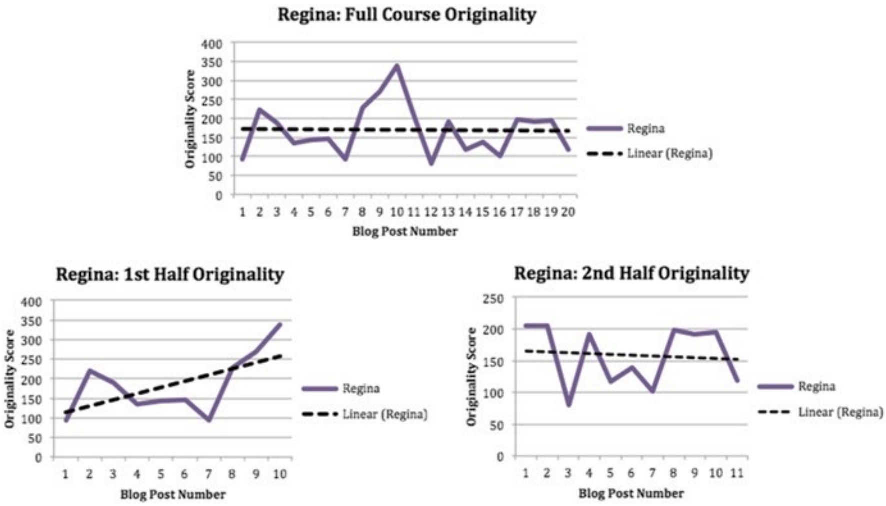

Fig. 8.6 Three originality slopes for Regina

students. In the case of Samantha, not only the slope for the second half of the course but in fact across every interval is positive. For Regina, however, an initial level of heightened originality drops off precipitously and never again recovers. In the previous section, Regina's results had appeared to be something of an outlier: her peak of originality, in Week 10, was significantly greater than the average class

originality. Whatever this one week's performance denotes, the results here would suggest that Regina's success was unsustainable. After that week, her originality results tail off significantly; not only her originality slope over the second half of the course, but indeed for the entire semester *overall* is negative. We can imagine different distributions of these data points. Indeed, one of the more prevalent trends among poorly performing students is a simply static slope, whereby there's no significant change from course beginning to course end.

One reason for these findings became clearer following a closer examination of the way in which the course itself had been designed. As the NPSY course syllabus had been established, the first half of the course was given over to discussing the elementary components of neurobiology: the anatomy of the brain, the composition of the nervous system, how neurons fire, and how neurotransmitters function. Any student who had taken an introductory biology course could have possessed a fairly adequate grasp of some portion of this material, and any student pursuing a more advanced course in neuropsychology could surely be counted upon to understand how neurons work. Only in the later weeks of the course did discussion become more abstract, touching upon topics such as "attention," "learning," and "memory." Here, conceivably, is where the course would become most difficult, and where all the material from the first half of the course would need to be applied to a variety of different contexts.

Although the Point of Originality was (originally) developed as a tool for intervention rather than a way of formally characterizing what successful learning looks like, one thing that these findings suggest is that when used with the full range of query terms that a course might cover, the Point of Originality tool might in fact provide evidence of *continuing originality* or, put more audaciously, of sustained learning. For as students like Samantha here worked their way further and further into the course, what we found is that they only seemed to produce an ever mounting engagement with the course materials. Based on the "bulk" nature of the query term list, what these values would need to suggest would be an increased and *increasing* capacity to deploy course concepts in more refined and more elaborate combinations. Exploring this aspect of the Point of Originality's potential will be a key interest of future research.

The Relationship Between Verbosity and Originality Scores

The large number of query terms used for this case study provides an opportunity to address a frequently encountered question regarding the Point of Originality's reliance on largely variable inputs like student writings. As with any analysis method that depends upon the summation of discrete values—even discrete *meaningful* values—one can imagine a case where the algorithm would produce an abundance of hits simply because a post *itself* was longer than usual. Such a scenario would produce a possible "false positive," whereby a student would either intentionally or inadvertently have "gamed the system," appearing to have attained mastery of course material merely by being more verbose.

The likelihood that such a scenario would occur, however, is theoretically remote. The inherent complexity of the lexical relationship between words would make it almost impossible for any individual to reproduce a body of text that would attract consistently high originality values from a number of different query terms. A long and largely mindless string of words ("neuroscience, brain, brain, brain, hypothalamus") would be unlikely to trigger a false positive because none of the most generic phrases would be sufficiently distant from the original query terms to generate a truly large originality score. With a sufficiently large list of query terms, as in the case study at hand, reproducing these conditions should become even more difficult, since the highest possible originality score would need to speak equally to dozens of different terms at once. There's a kind of built-in pedagogical paradox here, for if a student were to understand which terms *were* liable to appeal to every aspect of the course material, they would conceivably need to *already* have a fluent comprehension of the course content.

To determine the possibility of such a scenario, we nevertheless conducted a very simple evaluation of the degree to which the length of writings impacted accrued originality scores. For each of the students, we computed the ratio between the number of words that produced any manner of originality score (i.e., that produced a "hit") and how many words appeared in the writing sample (excluding stopwords).

$$\text{Ratio} = \frac{\left(\text{No. words in samples} - \text{Stopwords}\right)}{\text{No. words producing hits in samples}}$$

This ratio measures the degree to which the writing sample addresses all possible aspects of course content as indicated by the instructor's list of query terms. For this case study, the query term list encompasses all 112 unique query terms.

In a hypothetical case where a post consisted of five words that produced five hits, the ratio would be 5:5 or 1 (the quotient of the two variables). If the same post only produced four hits, the ratio would be 5:4 or 1.25. That same post producing three hits would yield a ratio of 5:3 or 1.67, and so on. If five hits were found in a post with ten words, the ratio would be 10:5 or 2.

Therefore, the fewer the words that garner hits, the higher the ratio. If a writing sample consistently produced hits relative to its length, the ratio would remain close to 1.

The scatterplot, shown in Fig. 8.7, shows the degree to which a relationship between the ratios and writing length exists within the dataset.

The *x*-axis of the plot shows, per student, the "hit ratio"—a ratio of the average number of words that produced a hit in a students' writing relative to the number of words the posts contained on average. The *y*-axis shows the "writing length," or average word count, for each student in the class. The data points are black circles if the particular student's overall average originality score was below the class average and shown as grey triangles if the score was above the class average.

Conducting a simple linear regression test on the distribution in the plot shows that there is a positive relationship between the two variables. In other words, as word count increased, the hit ratio decreased: those students who tended to produce more words also had fewer words producing hits for each new word.

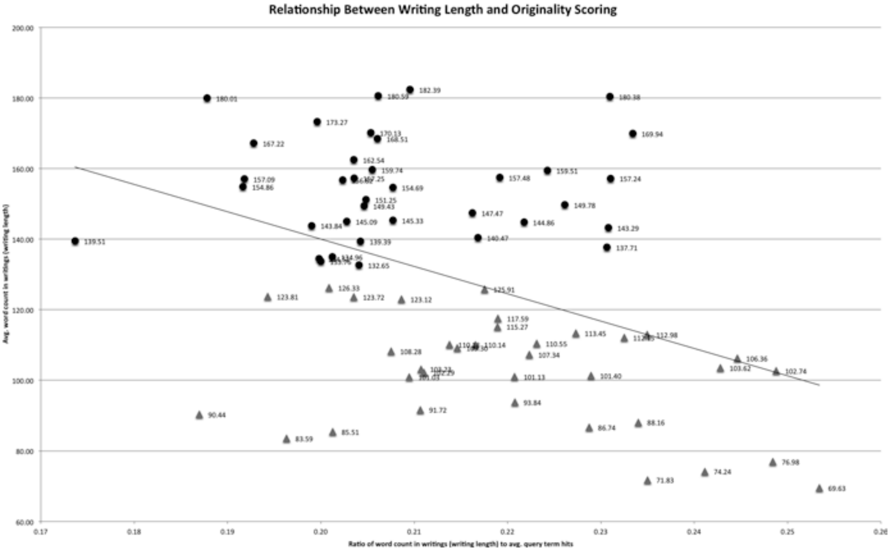

Fig. 8.7 Relationship between writing length and originality score

This evidence directly contradicts the supposition that the analysis would unfairly reward longer writings simply because they contained more words. In fact, the actual relationship would appear to be quite the opposite. Longer writings *can* receive higher originality scores, but simply producing additional text does not dramatically affect the system. Blue and red dots are distributed at a number of different points along the x-axis.

A Pearson correlation coefficient was computed to assess the relationship between the two variables. There was a strong positive correlation between the two variables. In other words, increases in writing length had students receive fewer hits of originality in their writings.

$$r(69) = 0.61, p < 0.01$$

Characterization of Vocabulary Sizes of Two Groups of Students

One of the immediate merits of the Point of Originality tool is its ability to quantify seemingly ephemeral aspects of language use and linguistic sophistication. The diagnostic predictions provided by the tool are primarily meant to provide instructors with insight into *when* and *to what extent* students are beginning to think comprehensively through course material. The large number of query terms used for this phase of the study, however, possibly allows us, as researchers, to look more closely into just *why* students have produced the originality values that they have, isolating what the relative vocabularies of individual students might look like, and suggesting what features of student writing correspond to higher originality values and, by

extension, likely greater achievement. Put differently, whereas the Point of Originality tool is primarily about the *students' progress* from week to week, our work as researchers permits us to identify the elements of *student writing* that directly lead to more favorable outcomes.

Since the query term list here was meant, in theory, to encompass every key concept discussed in the course, we generated a list of every word within every student's blog posts that had produced any matches with any query terms whatsoever. This effectively would have accounted for every word or phrase or topic that students could have produced that would have been germane to the course material. We then attempted to characterize the relative vocabularies for two groups of students: first, for the same nine students who had above received the highest final grade in the course, and likewise for the same seven students who had above received the lowest final grade in the course.

The most arresting immediate observation came simply from the scale of the relative vocabularies on display. Students in the highest performing group of students used a total of 7,390 unique words over their separate blog posts. By contrast, students in the lowest performing group of students used only 4,623 unique words. Those students who had done best in the course, in other words, had a vocabulary 59.85 % larger than those students who had performed the worst.

These findings might not be more than what might be reasonably intuited, but the magnitude of the difference between the two groups is nonetheless impressive. All students in the course were conceivably discussing the same course concepts, course concepts based on a highly specialized, specific, and scientific terminology. There are only so many ways to say "hypothalamus." Students who performed best in the course, however, *did* have the ability to stretch their vocabularies to encompass more varied appreciations of the concepts at hand. This is not to say that these students necessarily knew more words than their peers in absolute terms, but simply that they deployed a significantly larger number of terms that could still apply to the course material.

Results of this kind suggest a final implication for the Point of Originality tool. At one point, in an attempt to characterize the vocabularies of the highest and lowest performing students, we had considered placing the words used by each group into a "word cloud," which would display the words most commonly used by each group. Such a method of visualizing the data, however, proved inconclusive, if only because *both* groups of students (necessarily) had frequent recourse to the word "brain" or to the word "neuroscience." It is not the first, or second, or ninety-ninth most commonly used word that accounts for the difference in originality scores between students, but the hundredth, the thousandth, or the *six* thousandth most common word that the most successful students bring to bear but that the least successful students haven't even mentioned. Despite the finding of the section above ("The Relationship Between Verbosity and Originality Scores") that more successful students do not produce greater originality values by virtue of using *more* words, they do, finally, have more possible words to choose from, and it's the selection of those words, rather than others, that most accounts for the largest originality values.

This is one reason why the Point of Originality's ability to quickly compile, characterize, and visualize aspects of student writing might be preferred to other possible analysis methods. Insight into the applications and implications of student writing can only be accurately produced by a system that can account for the many features of what that writing might entail *holistically*. Such was the immediate effort of this phase of analysis: by accounting for as many query terms as possible, our goal was to guarantee the broadest possible application of the Point of Originality tool, ensuring that every student and every word was able, finally, to count.

8.6 Conclusions

Integrating technology into higher education curricula to extend the physical boundaries of the classroom can be of significant value, as it enables students to interact and learn outside of class time. This is particularly true in larger gateway courses, where there are fewer opportunities for students to engage in higher-order thinking and to construct their own understanding of core concepts. While the introduction of technology like blogging can create a successful learning experience, any large number of students creates additional noise that makes it harder for instructors to isolate the students most in need of help. This chapter described a method and tool with which student writing can be automatically analyzed to determine whether or not students have reached a *point of originality* in their writing, reflecting mastery of the course content. Through two case studies, higher point of originality values was shown to correspond strongly with likely student achievement.

Based on the two stages of research conducted in this study, the Point of Originality tool was able to provide an accurate and informed diagnostic of likely student success in a variety of course contexts, with a range of possible course exercises, and with different kinds of relevant terminology. Although initially intended to answer questions frequently raised regarding the tool, the requirements of our second stage of research, indeed, would appear to have only amplified its predictive possibilities. Increasing the number of query terms applied to student writing would make it easier for an instructor to generalize the results of what he or she had observed. Students identified as possibly struggling could now be understood to be struggling with aspects of the course as a whole, and broad trends in student application would be easier to intuitively spot. Even less time would now need to be spent understanding the results of the analysis, with more time, ideally, directed towards helping those students that the system had singled out.

The next step for the Point of Originality system is to modify the visualization method with these results firmly in mind. If all course query terms can be used to accurately gauge likely student success, then there's no longer any question, as there was in Phase I, of whether students might simply be more heavily invested in a different aspect of the course material. Under the framework established by Phase II, all students can be subject to the same analysis at the same time. If our original goal was to reduce the amount of time that would be required for instructors to

understand their students' progress from week to week, then presenting that information in a single window, through a single visualization, with just a single set of query terms, would only streamline the process further. Knowing what the Point of Originality tool is capable of, our goal, moving forward, is to better understand what instructors are capable of doing with the tool in place.

Acknowledgements Special thanks to the students in the CS 33b and NPSY 22 courses for providing the data for this research project and to Richard Alterman and Robert Sekuler, the instructors for those courses.

References

Alterman, R., & Larusson, J. A. (2009). Modeling participation within a community. In N. A. Taatgen & H. V. Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1680–1685). Austin, TX: Cognitive Science Society.

Alterman, R. & Larusson, J. A. (2010). Collaborative sensemaking in the blogosphere. In: *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1661–1666). Austin, TX: Cognitive Science Society.

Althaus, S. L. (1997). Computer-mediated communication in the university classroom: An experiment with on-line discussions. *Communication Education, 46*(3), 158–174.

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443–459). New York: Cambridge University Press.

Andriessen, J., Baker, M., & Suthers, D. (2003). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments*. New York: Springer.

Boyd-Graber, J., Blei, D., & Zhu, X. (2007). A topic model for word sense disambiguation. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1023–1033). Prague: Association for Computational Linguistics.

Brewer, C. A. (2004). Near real-time assessment of student learning and understanding in biology courses. *BioScience, 54*(11), 1034–1039.

Chi, M. T. H., & Van Lehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1*(1), 69–105.

Cooper, J. L., & Robinson, P. (2000). The argument for making large classes seem small. *New Directions for Teaching and Learning, 2000*(81), 5–16.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438–448.

Deitering, A., & Huston, S. (2004). Weblogs and the middle space for learning. *Academic Exchange Quarterly, 8*(4), 273–278.

Du, H. S., & Wagner, C. (2005). Learning with weblogs: An empirical investigation. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05)* (pp. 1–9). Washington, DC: IEEE Computer Society.

Duffy, P. (2008). Engaging the YouTube Google-eyed generation: Strategies for using Web 2.0 in Teaching and Learning. *The Electronic Journal of e-Learning, 6*, 119–130.

Eilam, B. (2002). Phases of learning: Ninth graders' skill acquisition. *Research in Science & Technological Education, 20*(1), 5–24.

Elias, T. (2011). *Learning analytics: Definitions, processes and potential*. Unpublished Internal Whitepaper of Athabasca University, Canada.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Ferdig, R. E., & Trammell, K. D. (2004). Content delivery in the "blogosphere". *Technological Horizons in Education, 31*(7), 12–16.

Flask (A Python Microframework). Retrieved from http://flask.pocoo.org

Fritz, J. (2010). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *The Internet and Higher Education, 14*(2), 89–97.

Gee, J. P., & Green, J. L. (1998). discourse analysis, learning, and social practice: A methodological study. *Review of Research in Education, 23*, 119–169.

Gerdeman, R. D., Russell, A. A., & Worden, K. J. (2007). Web-based student writing and reviewing in a large biology lecture course. *Journal of College Science Teaching, 36*(5), 7.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2–16.

Greer, L., & Heaney, P. J. (2004). Real-time analysis of student comprehension: An assessment of electronic student response technology in an introductory earth science course. *Journal of Geoscience Education, 52*, 345–351.

Hirst, T. (2011). Pragmatic analytics: Insight, representation, and structure. *Keynote Presented at the 1st International Conference Learning Analytics and Knowledge*. Banff, Alberta, Canada: ACM.

Huffaker, D. (2005). The educated blogger: Using weblogs to promote literacy in the classroom. *AACE Journal, 13*(2), 91–98.

Juang, Y. (2008). Learning by blogging: Warm-up and review lessons to facilitate knowledge building in classrooms. In: *Proceedings of the Eighth IEEE International Conference on Advanced Learning Technologies* (pp. 574–575). Washington, DC: IEEE.

Korthagen, F., & Lagerwerf, B. (1995). Levels in Learning. *Journal of Research in Science Teaching, 32*(10), 1011–1038.

Krasner, G. E., & Pope, S. T. (1988). A cookbook for using the model-view-controller user interface paradigm in the smalltalk-80 system. *Journal of Object Oriented Programming, 1*, 26–49.

Larusson, J. A., & Alterman, R. (2009). Wikis to support the "collaborative" part of collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 4*(4), 371–402.

Larusson, J. A., & White, B. (2012). Monitoring students' progress through their written point of originality. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK 2012)* (pp. 212–221). Vancouver, Canada: ACM.

Lebrun, M. (1999). *Des technologies pour enseigner et apprendre*. Brussels: De Boeck.

MacGregor, J. (2000). Restructuring large classes to create communities of learners. *New Directions for Teaching and Learning, 2000*(81), 47–61.

McAlpine, L., Weston, C., Berthiaume, D., Fairbank-Roch, G., & Owen, M. (2004). Reflection on teaching: Types and goals of reflection. *Educational Research and Evaluation: An International Journal on Theory and Practice, 10*(4), 337–363.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*(4), 693–720.

McKeachie, W. J., & Chism, N. (1986). *Teaching tips*. Washington, DC: DC Heath.

Mills-Jones, A. (1999). Active learning in IS education: Choosing effective strategies for teaching large classes in higher education. In: *Proceedings of 10th Australasian Conference on Information Systems* (pp. 1–3). ACIS.

Moore, M. T. (1985). The relationship between the originality of essays and variables in the problem-discovery process: A study of creative and noncreative middle school students. *Research in the Teaching of English, 19*(1), 84–95.

Nelson, N. (2001). Writing to learn: One theory, two rationales. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool: Integrating theory and practice* (pp. 23–36). Dordrecht, The Netherlands: Kluwer.

Nicol, D., & Boyle, J. (2003). Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education, 28*(4), 457–473.

Nikolova, S., Boyd-Graber, J., & Fellbaum, C. (2009). Collecting semantic similarity ratings to connect concepts in assistive communication tools. In A. Mehler, K. Kühnberger, H. Lobin, H.

Lüngen, A. Storrer, & A. Witt (Eds.), *Modelling, learning and processing of text-technological data structures. Springer studies in computational intelligence* (pp. 81–93). New York: Springer.

Oravec, J. A. (2002). Bookmarking the world: Weblog applications in education. *Journal of Adolescent & Adult Literacy, 45*(7), 616–621.

Python Programming Language—Official Website. Retrieved from http://python.org

Rashed, G., & Ahsan, R. (2012). Python in computational science: Applications and possibilities. *International Journal of Computer Applications, 26*(20), 26–30.

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*(2–3), 155–175.

Robinson, R. (2001). Calibrated Peer Review™ an application to increase student reading & writing skills. *The American Biology Teacher, 63*(7), 474–480.

Rosenshine, B., & Meister, C. (1995). Scaffolds for teaching higher-order cognitive strategies. In A. C. Ornstein (Ed.), *Teaching: Theory into practice* (pp. 134–153). Needham Heights, MA: Allyn & Bacon.

Shih, M. (1986). Content-based approaches to teaching academic writing. *TESOL Quarterly, 20*(4), 617–648.

Smith, M. L., & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal, 17*(4), 419–433.

SQLite Homepage. Retrieved from http://sqlite.org

Taylor, B. P. (1981). Content and written FORM: A two-way street. *TESOL Quarterly, 15*(1), 5–13.

Thomas, M. J. W. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning, 18*(3), 351–366.

Tynjälä, P., Mason, L., & Lonka, K. (2001). *Writing as a learning tool: Integrating theory and practice* (1st ed.). Dordrecht, The Netherlands: Kluwer.

Van Rossum, G., & Drake, F. L. (2003). *An introduction to python*. Bristol, UK: Network Theory Ltd.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19–27.

Williams, J. B., & Jacobs, J. (2004). Exploring the use of blogs as learning spaces in the higher education sector. *Australasian Journal of Educational Technology, 20*(2), 232–247.

Wise, L. (2005). *Blogs versus discussion forums in postgraduate online continuing medical education*. Paper presented at Blogtalk Down Under Conference, May 19-22, Sydney, Australia.

Yang, D., & Powers, D. M. (2005). Measuring semantic similarity in the taxonomy of WordNet. In: *Proceedings of the Twenty-Eighth Australasian Conference on Computer Science-Volume 38* (pp. 322–330). Darlinghurst, Australia: Australian Computer Society.

Index

J.A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*,
DOI 10.1007/978-1-4614-3305-7, © Springer Science+Business Media New York 2014